TRANSACTIONS
OF THE
AMERICAN INSTITUTE
OF
ELECTRICAL ENGINEERS

VOLUME 72
1953

PART I. COMMUNICATION AND ELECTRONICS

PUBLISHED BY THE
AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS
33 WEST 39TH STREET
NEW YORK 18, N. Y., U. S. A.

Copyright 1953 by the American Institute of Electrical Engineers
Printed in U. S. A. by The Mack Printing Company, Easton, Pa.

Communication and Electronics
Index for 1953

1. Technical Subject Index

A-C Network Analyzer, A New Design 60-Cycle. Davidson, Koll.........................141–5
A-C Null-Type Recorder with Balancing Amplifier Which Provides Damping and Suppresses the Quadrature Component. Williams, Payne.............................611–15; disc. 615
A-C Power Systems, Electronic Frequency Changer Used as Nonsynchronous Tie Between. Winograd.............................263–72; disc. 272
A-C Reactance on Voltage Regulation of 6-Phase Rectifiers, Influence of. Witzke, Kresser, Dillard.............................244–52; disc. 252
Accounting in the SATT System, Automatic Call Recording and. Ostline...................56–62
Accurate Radio-Frequency Microvoltages. Selby.............................158–63; disc. 163
Acoustic Impedance Measuring Techniques to Problems in Diathermy, Application of Electric and. Schwan, Carstensen........................106–10
Admittance Determinations of Highly Conducting Materials at Low Frequencies, Wheatstone Bridge. Schwan, Sittel.........................114–21
Air in Nonuniform Fields at Radio Frequencies, The Electric Strength of. Whitehead, Birx, Miller.....................................520–4
Air with Radium Isotopes, Production of Unipolar. Martin..................................771–6
Algebra to the Design of Electronic Switching Circuits, An Application of Boolean. Washburn.....380–8
AMA System, Principles of Tape-to-Card Conversion in the. Groth.................................42–52
AMA Toll Messages, Mechanized Billing of. Slade.............................175–82; disc. 182
Amplifier Which Provides Damping and Suppresses the Quadrature Component, A-C Null Type Recorder with Balancing. Williams, Payne.............................611–15; disc. 615
Amplifiers, Dielectric. Penney, Horsch, Sack...68–79
Analogue Shaft-Position Transducers, Digital-to-. O'Neil.................................37–41
Analogue Solution Applied to the Heat Conduction Problem in a Cartridge Fuse, An Analysis of an. Guile, Carne.................................861–7
Analysis of a Comb Filter Using Synchronously Commutated Capacitors. LePage, Cahn, Brown.....63–8
Analysis of a Series Circuit Containing a Nonlinear Capacitor, A Mathematical. Pipes...238–42; disc. 242
Analysis of an Analogue Solution Applied to the Heat Conduction Problem in a Cartridge Fuse, An. Guile, Carne.................................861–7
Analytical Operations, An Application of Magnetic Amplifier Circuits to Perform Multiplication and Other. Finzi, Mathias.......................455–61
Analyzer, A New Design 60-Cycle A-C Network. Davidson, Koll.............................141–5
Analyzer for Magnetic Amplifier Cores, A Permeability. Siskind.................................572–6
Analyzer for Magnetic Amplifiers, A Transient. Smith...461–5
Analyzer, The Solution of Sturm-Liouville Problems by D-C Network. Swenson.....................811–13
Analyzers for Pipe Network Analysis, The Use of Electric Network. Stephenson, Eaton.............857–61
Apparatus for NTSC Color Television, Transmitting Terminal. Burr.............................26–33
Application of Boolean Algebra to the Design of Electronic Switching Circuits, An. Washburn...380–8
Application of Electric and Acoustic Impedance Measuring Techniques to Problems in Diathermy. Schwan, Carstensen.........................106–10
Application of Magnetic Amplifier Circuits to Perform Multiplication and Other Analytical Operations, An. Finzi, Mathias.......................455–61
Applications of Integral Equations to the Solution of Nonlinear Electric Circuit Problems. Pipes...445–50
Applications of Recorded Machine Announcements, Telephone System. Bennett.............478–83
Asymmetry of the Resonance Curves, Effects of Harmonics on the Frequency of Oscillation as Well as on the. Mostafa...........................309–14
Automatic Call Recording and Accounting in the SATT System. Ostline.......................56–62
Automatic Computing Equipment, Solutions of Boundary-Value Problems on. Verzuh..........813–21
Automatic Control System with Provision for Scanning and Memory, An. Young.................392–5
Automatic Synchronizer, Cathode-Ray Synchroscope and. Ahmed..................217–22; disc. 222
Automatic Testing of Wired Relay Circuits. Hanson...850–7
Automatic Transfer Function Measuring and Recording System, An. Ehret, Hochschild, Embree, Grogan..............................664–9; disc. 669

B

Band-Pass Filters, Maximum Impedance Transformations in. O'Donnell, Williams..........110–13
Basic Concepts in the Analysis of Stationary Electric Circuits. Spence, Cahn........617–23; disc. 623
(Basic Sciences) A Capacitively Coupled Field Mapper for 2-Dimensional Distributed Source Field Problems. Gilbert, Gilbert...........345–9; disc. 349
(Basic Sciences) A Graphical Method for Flip-Flop Design. Johnston, Ratz.................52–60
(Basic Sciences) A Mathematical Analysis of a Series Circuit Containing a Nonlinear Capacitor. Pipes.............................238–42; disc. 242
(Basic Sciences) Analysis of a Comb Filter Using Synchronously Commutated Capacitors. LePage, Cahn, Brown.................................63–8
(Basic Sciences) Applications of Integral Equations to the Solution of Nonlinear Electric Circuit Problems. Pipes.................................445–50
(Basic Sciences) Block-Diagram Solutions for Vacuum-Tube Circuits. Stout.....................561–7
(Basic Sciences) Dielectric Breakdown of Sulfur Hexafluoride in Nonuniform Fields. Works, Dakin.............................682–7; disc. 687
(Basic Sciences) Electron-Ion Recombination at Low Pressures. Borowitz........................430–5
(Basic Sciences) Flow of Energy in D-C Machines. Hawthorne.................................438–45
(Basic Sciences) Fluorine-Containing Gaseous Dielectrics. Camilli, Plump....................93–102
(Basic Sciences) Mathematical Descriptions of Core Losses. Hale, Richardson................495–501
(Basic Sciences) New Method for the Optimum Design of Permanent Magnets Subjected to Demagnetizing Effects. Ziegler...............253–62; disc. 262
(Basic Sciences) Nondestructive Testing of Insulation. Brancato......................202–05; disc. 205
(Basic Sciences) Polyethylene Terephthalate—Its Use as a Capacitor Dielectric. Wooley, Kohman. McMahon.................................33–7
(Basic Sciences) Production of Unipolar Air with Radium Isotopes. Martin.....................771–6
(Basic Sciences) Propagation Mechanism of Impulse Corona and Breakdown in Oil. Liao, Anderson.............................641–7; disc. 647
(Basic Sciences) Some Fluorinated Liquid Dielectrics. Bashara.................................79–85
(Basic Sciences) Stressed Ferrites Having Rectangular Hysteresis Loops. Williams, Sherwood, Goertz, Schnettler.................................531–7
(Basic Sciences) Temperature Fields in Electrical Coils: Numerical Solutions. Schneider..........768–71
(Basic Sciences) The Effect of Minor Constituents in High Dielectric Constant Titanate Capacitors. Coffeen.................................704–09
(Basic Sciences) The Electric Strength of Air in Nonuniform Fields at Radio Frequencies. Whitehead, Birx, Miller.................................520–4
(Basic Sciences) The First-Order Behavior of Separable Oscillators. dePackh.....................450–5
(Basic Sciences) The Influence of a Transverse Magnetic Field on an Unconfined Glow Discharge. McBee, Dow.........................229–37; disc. 237
(Basic Sciences) The Permeability of Silicon-Iron at Very Low Flux Densities. Both..656–61; disc. 661
(Basic Sciences) The Rayleigh Method in Network Calculations. Schott, Heilfron.................89–93
Behavior of Separable Oscillators, The First-Order. dePackh.................................450–5
Billing of AMA Toll Messages, Mechanized. Slade.............................175–82; disc. 182
Block-Diagram Solutions for Vacuum-Tube Circuits. Stout.................................561–7
Boolean Algebra to the Design of Electronic Switching Circuits, An Application of. Washburn....380–8
Boundary-Value Problems on Automatic Computing Equipment, Solutions of. Verzuh.........813–21
Breakdown of Sulfur Hexafluoride in Nonuniform Fields, Dielectric. Works, Dakin........682–7; disc. 687
Bridge for Admittance Determinations of Highly Conducting Materials at Low Frequencies, Wheatstone. Schwan, Sittel.........................114–21
Bridge, Sensitivity and Output Formulas for the Resistance. Andrews...........213–16; disc. 216, 868
Broadcast Network, Development of Television Service Standards and Application to Design of a Television. Reed.........................838–50

C

Cable, A Nonarmored Submarine Telegraph. Lawton Hutchins.......................153–6; disc. 156
Cable in the Telegraph Industry, Wire and. Markley.............................298–308; disc. 308
Calculatiom, The Rayleigh Method in Network. Schott, Heilfron.............................89–93
Call Recording and Accounting in the SATT System, Automatic. Ostline.......................56–62
Capacitively Coupled Field Mapper for 2-Dimensional Distributed Source Field Problems, A. Gilbert, Gilbert.......................345–9; disc. 349
Capacitor, A Mathetmatical Analysis of a Series Circuit Containing a Nonlinear. Pipes..238–42; disc. 242
Capacitor Dielectric, Polyethylene Terephthalate—Its Use as a. Wooley, Kohman, McMahon.....33–7
Capacitors, Analysis of a Comb Filter Using Synchronously Commutated. LePage, Cahn, Brown..63–8
Capacitors in Power Systems with Rectifier Loads. Schmidt.................................14–17
Capacitors, The Effect of Minor Constituents in High Dielectric Constant Titanate. Coffeen....704–09
Card Conversion in the AMA System, Principles of Tape-to-. Groth.........................42–52
Carrier Systems, Co-ordination of M1 and N1 Telephone. Smith, Cornell, Jerome..........517–20
Carrier Systems, The Control of Noise and Crosstalk on N1. Aikens, Thaeler...........605–10; disc. 610
Carrier Technique, Railroad Communications Using High-Power. Hansen, Dean, Mullins.............................785–90; disc. 790
Carrier Telephone Deviation Regulator, Type-N. Perkins, Mahoney.........................757–62
Carrier Telephone, Program Transmission Over Type-N. Case, Kerney.........................791–5
Carrier Telephone, Type-ON. Fracassi, Kahl..713–21
Cartridge Fuse, An Analysis of an Analogue Solution Applied to the Heat Conduction Problem in a. Guile, Carne.................................861–7
Cathode—A Review, The Nickel Base Indirectly Heated Oxide. Bounds, Hambleton.............102–06
Cathode-Ray Synchroscope and Automatic Synchronizer. Ahmed.................217–22; disc. 222
Changer Used as Nonsynchronous Tie Between A-C Power Systems, Electronic Frequency. Winograd.............................263–72; disc. 272
Characteristics of FTL Type No. 20-B UHF Television Transmitter, Technical. Bradburd........555–61
Circuit Containing a Nonlinear Capacitor, A Mathematical Analysis of a Series. Pipes.............................238–42; disc. 242
Circuit Problems, Applications of Integral Equations to the Solution of Nonlinear Electric. Pipes...445–50
Circuitry Upon the Operating Hysteresis Loops, The Influence of Magnetic Amplifier. Lord.....721–8
Circuits, Automatic Testing of Wired Relay. Hanson...850–7
Circuits, Basic Concepts in the Analysis of Stationary Electric. Spence, Cahn........617–23; disc. 623
Circuits, Block-Diagram Solutions for Vacuum-Tube. Stout.................................561–7
Circuits Containing Iron-Cored Inductances, Oscillatory. Mostafa, El-Shishini.............274–83
Circuits in the 4A Toll Crossbar Switching System, Transistors and Their. Mallery..........388–92
Circuits, The Diode Matrix as a Component in Relay Switching. Bush.........................833–8
Circuits, The Map Method for Synthesis of Combinational Logic. Karnaugh......593–8; disc. 598
Circuits to Perform Multiplication and Other Analytical Operations, An Application of Magnetic Amplifier. Finzi, Mathias.......................455–61
Civil Defense, Communications for. Armstrong...315–26
Coast Area, Microwave as Applied to Railroad Operation in the Gulf. Thomas................800–04

Coaxial System—Amplifiers, L3. Morris, Lovell Dickinson.....505–17
Coaxial System—Equalization and Regulation, L3. Ketchledge, Finch.....414–30
Coaxial System—Television Terminals, L3. Rieke, Graham.....541–51
Coaxial System, The L3. Elmendorf, Ehrbar, Klie, Grossman.....395–413
Code Digital Quantizer, A Progressive. Raasch.....567–71
Coils: Numerical Solutions, Temperature Fields in Electrical. Schneider.....768–71
Color Kinescope, The RCA 3-Gun Shadow-Mask. Law.....134–40
Color Television, Transmitting Terminal Apparatus for NTSC. Burr.....26–33
Comb Filter Using Synchronously Commutated Capacitors, Analysis of a. LePage, Cahn, Brown.....63–8
Combinational Logic Circuits, The Map Method for Synthesis of, Karnaugh.....593–8; disc. 598
(Committee Report) Water Cooling Systems of Mercury-Arc Rectifiers. (53–288).....465–75
(Communication) A Subscriber Toll Dialing Tape Reader. Blashfield.....17–21
(Communication) Automatic Call Recording and Accounting in the SATT System. Ostline.....56–62
(Communication) Automatic Testing of Wired Relay Circuits. Hanson.....850–7
(Communication) Co-ordination of M1 and N1 Telephone Carrier Systems. Smith, Cornell, Jerome.....517–20
Communication Facilities Serving Power Stations and Substations, Protection of Wire. Alexander.....587–91; disc. 591
(Communication) Heat Dissipation from Toll Transmission Equipment. Coy.....762–8
(Communication) L3 Coaxial System—Amplifiers. Morris, Lovell, Dickinson.....505–17
(Communication) L3 Coaxial System—Equalization and Regulation. Ketchledge, Finch.....414–30
(Communication) Mechanized Billing of AMA Toll Messages. Slade.....175–82; disc. 182
(Communication) Principles of Tape-to-Card Conversion in the AMA System. Groth.....42–52
Communication System, The New Jersey Turnpike—A Unique Highway. Godley, Neubauer, Marsh.....360–9
Communication Systems, A Miniature Compandor for General Use in Wire and Radio. Boxall, Caruthers.....804–11
(Communication) Telephone-System Applications of Recorded Machine Announcements. Bennett.....478–83
(Communication) The Control of Noise and Crosstalk on N1 Carrier Systems. Aikens, Thaeler.....605–10; disc. 610
(Communication) The Diode Matrix as a Component in Relay Switching Circuits. Bush.....833–8
(Communication) The L3 Coaxial System. Elmendorf. Ehrbar, Klie, Grossman.....395–413
(Communication) The Map Method for Synthesis of Combinational Logic Circuits. Karnaugh.....593–8; disc. 598
(Communication) Transistors and Their Circuits in the 4A Toll Crossbar Switching System. Mallery.....388–92
(Communication) Transmission Design of Intertoll Telephone Trunks. Huntley.....670–6; disc. 676
Communication Transmitter, A Million-Watt Naval. Walter.....369–74
(Communication) Type-N Carrier Telephone Deviation Regulator. Perkins, Mahoney.....757–62
(Communication) Type-7 Crossbar Selector. Hutton.....830–3
Communications for Civil Defense. Armstrong.....315–26
(Communications) Program Transmission Over Type-N Carrier Telephone. Case, Kerney.....791–5
(Communications) Type-ON Carrier Telephone. Fracassi, Kahl.....713–21
(Communications) Use of Radio and Radar in the Petroleum Industry. Rust.....796–800
Communications Using High-Power Carrier Technique, Railroad. Hansen, Dean, Mullins.....785–90; disc. 790
Commutated Capacitors, Analysis of a Comb Filter Using Synchronously. LePage, Cahn, Brown.....63–8
Commutatorless D-C Tachometer, A. Eckels, Peck.....625–8; disc. 628
Compandor for General Use in Wire and Radio Communication Systems, A Miniature. Boxall, Caruthers.....804–11
Compensator for Telephotograph Systems, A Level. Jones, Phelps.....537–41
Component in Relay Switching Circuits, The Diode Matrix as a. Bush.....833–8
Computation of Impedance and Efficiency of Transmission Line with High Standing-Wave Ratio. Macalpine.....334–9
(Computing Devices) A New Design 60-Cycle A-C Network Analyzer. Davidson, Koll.....141–5
(Computing Devices) A Progressive Code Digital Quantizer. Raasch.....567–71
(Computing Devices) An Analysis of an Analogue Solution Applied to the Heat Conduction Problem in a Cartridge Fuse. Guile, Carne.....861–7
(Computing Devices) Digital-to-Analogue Shaft-Position Transducers. O'Neil.....37–41
(Computing Devices) Nondestructive Sensing of Magnetic Cores. Buck, Frank.....822–30
(Computing Devices) Pulse Response Characteristics of Rectangular-Hysteresis-Loop Ferromagnetic Materials. Wylen.....648–55; disc. 655
(Computing Devices) The Solution of Sturm-Liouville Problems by D-C Network Analyzer. Swenson.....811–13
(Computing Devices) The Use of Electric Network Analyzers for Pipe Network Analysis. Stephenson, Eaton.....857–61
Computing Equipment, Solutions of Boundary-Value Problems on Automatic. Verzuh.....813–21
Concepts in the Analysis of Stationary Electric Circuits, Basic. Spence, Cahn.....617–23; disc. 623
Concepts in the Mid-Frequency Range, Electron Tube Voltmeter. Hoadley.....501–04
Conduction Problem in a Cartridge Fuse, An Analysis of an Analogue Solution Applied to the Heat. Guile, Carne.....861–7
Conductive Plastic, Forcing Function Generator Employing. Norman.....576–81
Constants for Rectifiers, Equivalent Machine. Dortort.....435–8
Constituents in High Dielectric Constant Titanate Capacitors, The Effect of Minor. Coffeen.....704–09
Construction of Shielded Rooms, The Use of Steel Sheet for the. Intrator.....599–604; disc. 604
Control Considerations, Nuclear Power Plant. Schultz.....22–6
Control of Noise and Crosstalk on N1 Carrier Systems, The. Aikens, Thaeler.....605–10; disc. 610
Control System with Provision for Scanning and Memory, An Automatic. Young.....392–5
Conventions, Television Coverage of the National Political. Ralston, Wickline.....1–14
Cooling Systems of Mercury-Arc Rectifiers, Water. (Committee Report).....465–75
Co-ordination of M1 and N1 Telephone Carrier Systems. Smith, Cornell, Jerome.....517–20
Copper Temperature Indicator, Generator Stator. Brownlee, Brown.....676–8; disc. 678
Core Losses, Mathematical Descriptions of. Hale, Richardson.....495–501
Core Materials, An Instability of Self-Saturating Magnetic Amplifiers Using Rectangular Loop. Batdorf. Johnson.....223–7; disc. 227
Core Materials Employed in Magnetic Amplifiers, Dynamic Hysteresis Loops of Several. Lord.....85–8
Cores, A Permeability Analyzer for Magnetic Amplifier. Siskind.....572–6
Cores, Nondestructive Sensing of Magnetic. Buck, Frank.....822–30
Corona and Breakdown in Oil, Propagation Mechanism of Impulse. Liao, Anderson.....641–7; disc. 647
Correction of Frequency Errors in Wattmeters. Freeman.....679–81; disc. 681
Crossbar Selector, Type-7. Hutton.....830–3
Crossbar Switching System, Transistors and Their Circuits in the 4A Toll. Mallery.....388–92
Crosstalk on N1 Carrier Systems, The Control of Noise and. Aikens, Thaeler.....605–10; disc. 610
Current Standard, Preliminary Development of a Magnettor. Felch, Potter.....524–3
Curves, Effects of Harmonics on the Frequency of Oscillation as Well as on the Asymmetry of the Resonance. Mostafa.....309–14
Curves for 6-Phase Double-Way and Double-Wye Rectifiers, Extended Regulation. Dortort.....192–8; disc. 198

D

D-C Load, Saturable Reactors with Inductive. Storm: Part II. Transient Response.....182–90; disc. 191
D-C Machines, Flow of Energy in. Hawthorne.....438–45
D-C Network Analyzer, The Solution of Sturm-Liouville Problems by. Swenson.....811–13
D-C Tachometer, A Commutatorless. Eckels, Peck.....625–8; disc. 628
Damping and Suppresses the Quadrature Component, A-C Null-Type Recorder with Balancing Amplifier Which Provides. Williams, Payne.....611–15; disc. 615
Defense, Communications for Civil. Armstrong.....315–26
Densities, The Permeability of Silicon-Iron at Very Low Flux. Both.....656–61; disc. 661
Descriptions of Core Losses, Mathematical. Hale, Richardson.....495–501
Design, A Graphical Method for Flip-Flop. Johnston, Ratz.....52–60
Design of Permanent Magnets Subjected to Demagnetizing Effects, New Method for the Optimum. Ziegler.....253–62; disc. 262
Design 60-Cycle A-C Network Analyzer, A New. Davidson, Koll.....141–5
Detectors, A New Method for Treating Electron Tubes when Used as Superregenerative. Mostafa, El-Shishini:
  Part I. Superregenerative Circuits Under No-Signal Condition.....207–13
  Part II. Superregenerative Circuits Under Signal Conditions.....283–9
  Part III. Experimental Investigation of Superregenerative Circuits.....290–7
Determinations of Highly Conducting Materials at Low Frequencies, Wheatstone Bridge for Admittance. Schwan, Sittel.....114–21
Development of a Magnettor Current Standard, Preliminary. Felch, Potter.....524–31
Development of Television Service Standards and Application to Design of a Television Broadcast Network. Reed.....838–50
Deviation Regulator, Type-N Carrier Telephone. Perkins, Mahoney.....757–62
Diagram Solutions for Vacuum-Tube Circuits, Block-. Stout.....561–7
Dialing Tape Reader, A Subscriber Toll. Blashfield.....17–21
Diathermy, Application of Electric and Acoustic Impedance Measuring Techniques to Problems in. Schwan, Carstensen.....106–10
Diathermy, Heating of Fat-Muscle Layers by Electromagnetic and Ultrasonic. Schwan, Carstensen, Li.....483–8
Dielectric Amplifiers. Penney, Horsch, Sack.....68–79
Dielectric Breakdown of Sulfur Hexafluoride in Nonuniform Fields. Works, Dakin.....682–7; disc. 687
Dielectric Constant Titanate Capacitors, The Effect of Minor Constituents in High. Coffeen.....704–09
Dielectric, Polyethylene Terephthalate—Its Use as a Capacitor. Wooley, Kohman, McMahon.....33–7
Dielectrics, Fluorine-Containing Gaseous. Camilli, Plump.....93–102
Dielectrics, Some Fluorinated Liquid. Bashara.....79–85
Digital Quantizer, A Progressive Code. Raasch.....567–71
Digital-to-Analogue Shaft-Position Transducers. O'Neil.....37–41
Diode Matrix as a Component in Relay Switching Circuits, The. Bush.....833–8
Discharge, The Influence of a Transverse Magnetic Field on an Unconfined Glow. McBee, Dow.....229–37; disc. 237
Dissipation from Toll Transmission Equipment, Heat. Coy.....762–8
Double-Wye Rectifiers, Extended Regulation Curves for 6-Phase Double-Way and. Dortort.....192–8; disc. 198
Dynamic Hysteresis Loops of Several Core Materials Employed in Magnetic Amplifiers. Lord.....85–8

E

Effect of Minor Constituents in High Dielectric Constant Titanate Capacitors, The. Coffeen.....704–09
Effects of Harmonics on the Frequency of Oscillation as Well as on the Asymmetry of the Resonance Curves Mostafa.....309–14
Efficiency of Transmission Line with High Standing-Wave Ratio, Computation of Impedance and. Macalpine.....334–9
Electric and Acoustic Impedance Measuring Techniques to Problems in Diathermy, Application of. Schwan, Carstensen.....106–10
Electric Circuit Problems, Applications of Integral Equations to the Solution of Nonlinear. Pipes.....445–50
Electric Circuits, Basic Concepts in the Analysis of Stationary. Spence, Cahn.....617–23; disc. 623
Electric Strength of Air in Nonuniform Fields at Radio Frequencies, The. Whitehead, Birx, Miller.....520–4
Electrical Coils: Numerical Solutions, Temperature Fields in. Schneider.....768–71
Electrocardiographic Theory, The Elements of. Frank.....125–34
Electromagnetic and Ultrasonic Diathermy, Heating of Fat-Muscle Layers by. Schwan, Carstensen, Li.....483–8
Electron-Ion Recombination at Low Pressures. Borowitz.....430–5
Electron Tube Voltmeter Concepts in the Mid-Frequency Range. Hoadley.....501–04
Electron Tubes at High Altitudes and High Ambient Temperatures, Heat Transfer from. Buckland.....698–703; disc. 703
Electron Tubes when Used as Superregenerative Detectors, A New Method for Treating. Mostafa, El-Shishini:
  Part I. Superregenerative Circuits Under No-Signal Condition.....207–13

Part II. Superregenerative Circuits Under Signal Conditions....283–9
Part III. Experimental Investigation of Superregenerative Circuits....290–7
Electronic Equipment, Reliability in Industrial. Cook....351–60
Electronic Frequency Changer Used as Nonsynchronous Tie Between A-C Power Systems. Winograd....263–72; disc. 272
Electronic Switching Circuits, An Application of Boolean Algebra to the Design of. Washburn....380–8
(Electronics) Gas Tube Filament Connections. Brenner, Schmitt....776–84; disc. 784
(Electronics) High-Power Industrial Vacuum Tubes Having Thoriated-Tungsten Filaments. Ayer....121–5
(Electronics) Water Cooling Systems of Mercury-Arc Rectifiers. (Committee Report)....465–75
Elements of Electrocardiographic Theory, The. Frank....125–34
Energy in D-C Machines, Flow of. Hawthorne....438–45
Equalization and Regulation, L3 Coaxial System—. Ketchledge, Finch....414–30
Equations to the Solution of Nonlinear Electric Circuit Problems, Applications of Integral. Pipes....445–50
Equipment for Million Watt Navy Transmitter, Frequency Generating. Robertson....475–7
Equipment, Heat Dissipation from Toll Transmission. Coy....762–8
Equipment, Reliability in Industrial Electronic. Cook....351–60
Equivalent Machine Constants for Rectifiers. Dortort....435–8
Errors in Wattmeters, Correction of Frequency. Freeman....679–81; disc. 681
Extended Regulation Curves for 6-Phase Double-Way and Double-Wye Rectifiers. Dortort....192–8; disc. 198

F

FTL Type No. 20-B UHF Television Transmitter Technical Characteristics of. Bradburd....555–61
Fast Response with Magnetic Amplifiers. Scorgie....741–9
Fat-Muscle Layers by Electromagnetic and Ultrasonic Diathermy, Heating of. Schwan, Carstensen, Li....483–8
Ferrites Having Rectangular Hysteresis Loops, Stressed. Williams, Sherwood, Goertz, Schnettler....531–7
Ferromagnetic Materials, Pulse Response Characteristics of Rectangular-Hysteresis-Loop. Wylen....648–55; disc. 655
Field Mapper for 2-Dimensional Distributed Source Field Problems, A Capacitively Coupled. Gilbert, Gilbert....345–9; disc. 349
Field on an Unconfined Glow Discharge, The Influence of a Transverse Magnetic. McBee, Dow....229–37; disc. 237
Fields in Electrical Coils: Numerical Solutions, Temperature. Schneider....768–71
Filament Connections, Gas Tube. Brenner, Schmitt....776–84; disc. 784
Filter Using Synchronously Commutated Capacitors, Analysis of a Comb. LePage, Cahn, Brown....63–8
Filters, Maximum Impedance Transformations in Band-Pass. O'Donnell, Williams....110–13
First-Order Behavior of Separable Oscillators, The. dePackh....450–5
Flip-Flop Design, A Graphical Method for. Johnston, Ratz....52–60
Flow of Energy in D-C Machines. Hawthorne....438–45
Fluorinated Liquid Dielectrics, Some. Bashara....79–85
Fluorine-Containing Gaseous Dielectrics. Camilli, Plump....93–102
Flux Densities, The Permeability of Silicon-Iron at Very Low. Both....656–61; disc. 661
Flux Preset High-Speed Magnetic Amplifiers. House....728–35
Forcing Function Generator Employing Conductive Plastic. Norman....576–81
Formulas for the Resistance Bridge, Sensitivity and Output. Andress....213–16; disc. 216, 868
40- to 4,000-Microwatt Power Meter. Lange....492–4
Frequency Changer Used as Nonsynchronous Tie Between A-C Power Systems, Electronic. Winograd....263–72; disc. 272
Frequency Errors in Wattmeters, Correction of. Freeman....679–81; disc. 681
Frequency Generating Equipment for Million Watt Navy Transmitter. Robertson....475–7
Frequency Microvoltages, Accurate Radio-. Selby....158–63; disc. 163
Frequency of Oscillation as Well as on the Asymmetry of the Resonance Curves, Effects of Harmonics on The. Mostafa....309–14
Fuse, An Analysis of an Analogue Solution Applied to the Heat Conduction Problem in a Cartridge Fuse. Guile, Carne....861–7

G

Gas Tube Filament Connections. Brenner, Schmitt....776–84; disc. 784
Gaseous Dielectrics, Fluorine-Containing. Camilli, Plump....93–102
Generating Equipment for Million Watt Navy Transmitter, Frequency. Robertson....475–7
Generator Employing Conductive Plastic, Forcing Function. Norman....576–81
Generator Stator Copper Temperature Indicator. Brownlee, Brown....676–8; disc. 678
Glow Discharge, The Influence of a Transverse Magnetic Field on an Unconfined. McBee, Dow....229–37; disc. 237
Graphical Method for Flip-Flop Design, A. Johnston, Ratz....52–60
Gulf Coast Area, Microwave as Applied to Railroad Operation in The. Thomas....800–04

H

Harmonics on the Frequency of Oscillation as Well as on the Asymmetry of the Resonance Curves, Effects of. Mostafa....309–14
Heat Conduction Problem in a Cartridge Fuse, An Analysis of an Analogue Solution Applied to the. Guile, Carne....861–7
Heat Dissipation from Toll Transmission Equipment. Coy....762–8
Heat Transfer from Electron Tubes at High Altitudes and High Ambient Temperatures. Buckland....698–703; disc. 703
Heated Oxide Cathode—A Review, The Nickel Base Indirectly. Bounds, Hambleton....102–06
Heating of Fat-Muscle Layers by Electromagnetic and Ultrasonic Diathermy. Schwan, Carstensen, Li....483–8
Hexafluoride in Nonuniform Fields, Dielectric Breakdown of Sulfur. Works, Dakin....682–7; disc. 687
High Altitudes and High Ambient Temperatures, Heat Transfer from Electron Tubes at. Buckland....698–703; disc. 703
High-Power Carrier Technique, Railroad Communications Using. Hansen, Dean, Mullins....785–90; disc. 790
High-Power Industrial Vacuum Tubes Having Thoriated-Tungsten Filaments. Ayer....121–5
High-Speed Magnetic Amplifiers, Flux Preset. House....728–35
Highway Communication System, The New Jersey Turnpike—A Unique. Godley, Neubauer, Marsh....360–9
Hysteresis-Loop Ferromagnetic Materials, Pulse Response Characteristics of Rectangular-. Wylen....648–55; disc. 655
Hysteresis Loops of Several Core Materials Employed in Magnetic Amplifiers, Dynamic. Lord....85–8
Hysteresis Loops, Stressed Ferrites Having Rectangular. Williams, Sherwood, Goertz, Schnettler....531–7
Hysteresis Loops, The Influence of Magnetic Amplifier Circuitry Upon the Operating. Lord....721–8

I

Ignitron Rectifier Installation, Load-Dropping Tests on a Large. Pope, Dillard, Marcum....164–72; disc. 172
Impedance and Efficiency of Transmission Line with High Standing-Wave Ratio, Computation of. Macalpine....334–9
Impedance Measuring Techniques to Problems in Diathermy, Application of Electric and Acoustic. Schwan, Carstensen....106–10
Impedance Transformations in Band-Pass Filters, Maximum. O'Donnell, Williams....110–13
Impulse Corona and Breakdown in Oil, Propagation Mechanism of. Liao, Anderson....641–7; disc. 647
Indicator, Generator Stator Copper Temperature. Brownlee, Brown....676–8; disc. 678
Inductances, Oscillatory Circuits Containing Iron-Cored. Mostafa, El-Shishini....274–83
Inductive D-C Load, Saturable Reactors with. Storm: Part II. Transient Response....182–90; disc. 191
Industrial Electronic Equipment, Reliability in. Cook....351–60
Industrial Vacuum Tubes Having Thoriated-Tungsten Filaments, High-Power. Ayer....121–5
Industry, Use of Radio and Radar in the Petroleum. Rust....796–800
Industry, Wire and Cable in the Telegraph. Markley....298–308; disc. 308
Influence of a Transverse Magnetic Field on an Unconfined Glow Discharge, The. McBee, Dow....229–37; disc. 237
Influence of A-C Reactance on Voltage Regulation of 6-Phase Rectifiers. Witzke, Kresser, Dillard....244–52; disc. 252
Influence of Magnetic Amplifiers Circuitry Upon the Operating Hysteresis Loops, The. Lord....721–8
Instability of Self-Saturating Magnetic Amplifiers Using Rectangular Loop Core Materials, An. Batdorf. Johnson....223–7; disc. 227
Installation, Load-Dropping Tests on a Large Ignitron Rectifier. Pope, Dillard, Marcum....164–72; disc. 172
(Instruments) A Commutatorless D-C Tachometer. Eckels, Peck....625–8; disc. 628
(Instruments) A Permeability Analyzer for Magnetic Amplifier Cores. Siskind....572–6
(Instruments) Correction of Frequency Errors in Wattmeters. Freeman....679–81; disc. 681
Instruments for Simplified Quality-Control Measurements, Integrating. Longenecker....375–80
(Instruments) Generator Stator Copper Temperature Indicator. Brownlee, Brown....676–8; disc. 678
(Instruments) Preliminary Development of a Magnettor Current Standard. Felch, Potter....524–31
Insulation, Nondestructive Testing of. Brancato....202–05; disc. 205
Insulation Resistance, A Universal Meter for Measuring Voltages at High Impedances, Micromicroamperes, and. Clark, Watson, Mergner....551–5
Integral Equations to the Solution of Nonlinear Electric Circuit Problems, Applications of. Pipes....445–50
Integrating Instruments for Simplified Quality-Control Measurements. Longenecker....375–80
Interference, The Measurement of Random Monochrome Video. Barstow, Christopher....735–41
Intertoll Telephone Trunks, Transmission Design of. Huntley....670–6; disc. 676
Ion Recombination at Low Pressures, Electron-. Borowitz....430–5
Iron at Very Low Flux Densities, The Permeability of Silicon-. Both....656–61; disc. 661
Iron-Cored Inductances, Oscillatory Circuits Containing. Mostafa, El-Shishini....274–83
Isotopes, Production of Unipolar Air with Radium. Martin....771–6

J

Joint Use of Wood Pole Lines with Increasing Line Voltage, Some Aspects of. Bloecker, Bullard....709–12

K

Kinescope, The RCA 3-Gun Shadow-Mask Color. Law....134–40

L

L3 Coaxial System—Amplifiers. Morris, Lovell, Dickinson....505–17
L3 Coaxial System—Equalization and Regulation Ketchledge, Finch....414–30
L3 Coaxial System—Television Terminals. Rieke, Graham....541–51
L3 Coaxial System, The. Elmendorf, Ehrbar, Klie, Grossman....395–413
Layers by Electromagnetic and Ultrasonic Diathermy, Heating of Fat-Muscle. Schwan, Carstensen, Li....483–8
Level Compensator for Telephotograph Systems, A. Jones, Phelps....537–41
Lines with Increasing Line Voltage, Some Aspects of Joint Use of Wood Pole. Bloecker, Bullard....709–12
Liouville Problems by D-C Network Analyzer, The Solution of Sturm-. Swenson....811–13
Liquid Dielectrics, Some Fluorinated. Bashara....79–85
Load-Dropping Tests on a Large Ignitron Rectifier Installation. Pope, Dillard, Marcum....164–72; disc. 172
Loads, Capacitors in Power Systems with Rectifier. Schmidt....14–17
Logic Circuits, The Map Method for Synthesis of Combinational. Karnaugh....593–8; disc. 598
Loop Core Materials, An Instability of Self-Saturating Magnetic Amplifiers Using Rectangular. Batdorf. Johnson....223–7; disc. 227
Loops of Several Core Materials Employed in Magnetic Amplifiers, Dynamic Hysteresis. Lord....85–8
Loops, Stressed Ferrites Having Rectangular Hysteresis. Williams, Sherwood, Goertz, Schnettler....531–7
Loops, The Influence of Magnetic Amplifier Circuitry Upon the Operating Hysteresis. Lord....721–8
Losses, Mathematical Descriptions of Core. Hale, Richardson....495–501
Low Flux Densities, The Permeability of Silicon-Iron at Very. Both....656–61; disc. 661
Low Pressures, Electron-Ion Recombination at. Borowitz....430–5

M

Ml and Nl Telephone Carrier Systems, Co-ordination of. Smith, Cornell, Jerome....517–20
Machine Announcements, Telephone-System Applications of Recorded. Bennett....478–83
Machine Constants for Rectifiers, Equivalent. Dortort....435–8
Machines, Flow of Energy in D-C. Hawthorne..438–45
Magnetic Amplifier Circuitry Upon the Operating Hysteresis Loops, The Influence of. Lord....721–8
Magnetic Amplifier Circuits to Perform Multiplication and Other Analytical Operations, an Application of. Finzi, Mathias....455–61
Magnetic Amplifier Cores, A Permeability Analyzer for. Siskind....572–6
Magnetic Amplifier for Synchros, A. Van Allen....749–56; disc. 756
Magnetic Amplifiers, A Transient Analyzer for. Smith....461–5
(Magnetic Amplifiers) Dielectric Amplifiers. Penney, Horsch, Sack....68–79
Magnetic Amplifiers, Dynamic Hysteresis Loops of Several Core Materials Employed in. Lord...85–8
Magnetic Amplifiers, Fast Response with. Scorgie....741–9
Magnetic Amplifiers, Flux Preset High-Speed. House....728–35
Magnetic Amplifiers Using Rectangular Loop Core Materials, An Instability of Self-Saturating. Batdorf, Johnson....223–7; disc. 227
Magnetic Amplifiers with Square-Loop Core Materials, Theory of. Storm....629–37; disc. 637
Magnetic Cores, Nondestructive Sensing of. Buck, Frank....822–30
Magnetic Field on an Unconfined Glow Discharge, The Influence of a Transverse. McBee, Dow....229–37; disc. 237
Magnetic Tape Recording, Synchronized. Ranger....581–6
Magnets Subjected to Demagnetizing Effects, New Method for the Optimum Design of Permanent. Ziegler....253–62; disc. 262
Magnettor Current Standard, Preliminary Development of a. Felch, Potter....524–31
Map Method for Synthesis of Combinational Logic Circuits, The. Karnaugh....593–8; disc. 598
Mapper for 2-Dimensional Distributed Source Field Problems, A Capacitively Coupled Field. Gilbert, Gilbert....345–9; disc. 349
Materials at Low Frequencies, Wheatstone Bridge for Admittance Determinations of Highly Conducting. Schwan, Sittel....114–21
Materials, Theory of Magnetic Amplifiers with Square-Loop Core. Storm....629–37; disc. 637
Mathematical Analysis of a Series Circuit Containing a Nonlinear Capacitor, A. Pipes....238–42; disc. 242
Mathematical Descriptions of Core Losses. Hale, Richardson....495–501
Matrix as a Component in Relay Switching Circuits, The Diode. Bush....833–8
Maximum Impedance Transformations in Band-Pass Filters. O'Donnell, Williams....110–13
Measurement of Random Monochrome Video Interference, The. Barstow, Christopher....735–41
Measurements, Integrating Instruments for Simplified Quality-Control. Longenecker....375–80
Measuring and Recording System, An Automatic Transfer Function. Ehret, Hochschild, Embree, Grogan....664–9; disc. 669
Mechanism of Impulse Corona and Breakdown in Oil, Propagation. Liao, Anderson....641–7; disc. 647
Mechanized Billing of AMA Toll Messages. Slade....175–82; disc. 182
(Medicine) The Elements of Electrocardiographic Theory. Frank....125–34
Memory, An Automatic Control System with Provision for Scanning and. Young....392–5
Mercury-Arc Rectifiers, Water Cooling Systems of. (Committee Report)....465–75
Meter for Measuring Voltages at High Impedances, Micromicroamperes, and Insulation Resistance, A Universal. Clark, Watson, Mergner....551–5
Meter, 40- to 4,000-Microwatt Power. Lange....492–4
Method for Flip-Flop Design, A Graphical. Johnston, Ratz....52–60
Method for the Optimum Design of Permanent Magnets Subjected to Demagnetizing Effects, New. Ziegler....253–62; disc. 262
Method for Treating Electron Tubes When Used as Superregenerative Detectors, A New. Mostafa, El-Shishini:
  Part I. Superregenerative Circuits Under No-Signal Condition....207–13
  Part II. Superregenerative Circuits Under Signal conditions....283–9
  Part III. Experimental Investigation of Superregenerative Circuits....290–7
Method in Network Calculations, The Rayleigh. Schott, Heilfron....89–93
Micromicroamperes, and Insulation Resistance, A Universal Meter for Measuring Voltages at High Impedances. Clark, Watson, Mergner....551–5
Microvoltages, Accurate Radio-Frequency. Selby....158–63; disc. 163
Microwatt Power Meter, 40- to 4,000-. Lange....492–4
Microwave as Applied to Railroad Operation in the Gulf Coast Area. Thomas....800–04
Microwave-Radio Routes, Path Testing for. Campbell....326–34
Mid-Frequency Range, Electron-Tube Voltmeter Concepts in the. Hoadley....501–04
Million-Watt Naval Communication Transmitter, A. Walter....369–74
Million Watt Navy Transmitter, Frequency Generating Equipment for. Robertson....475–7
Miniature Compandor for General Use in Wire and Radio Communication Systems, A. Boxall, Caruthers....804–11
Monochrome Video Interference, The Measurement of Random. Barstow, Christopher....735–41
Multiplication and Other Analytical Operations, An Application of Magnetic Amplifier Circuits to Perform. Finzi, Mathias....455–61

N

N1 Carrier Systems, The Control of Noise and Crosstalk on. Aikens, Thaeler....605–10; disc. 610
Nl Telephone Carrier Systems, Co-ordination of Ml and. Smith, Cornell, Jerome....517–20
National Political Conventions, Television Coverage of the. Ralston, Wickline....1–14
Naval Communication Transmitter, A. Million-Watt. Walter....369–74
Navy Transmitter, Frequency Generating Equipment for Million Watt. Robertson....475–7
Network Analyzer, A New Design 60-Cycle A-C. Davidson, Koll....141–5
Network Analyzer, The Solution of Sturm-Liouville Problems by D-C. Swenson....811–13
Network Analyzers for Pipe Network Analysis, The Use of Electric. Stephenson, Eaton....857–61
Network Calculations, The Rayleigh Method in. Schott, Heilfron....89–93
Network, Development of Television Service Standards and Application to Design of a Television Broadcast. Reed....838–50
New Jersey Turnpike—A Unique Highway Communication System, The. Godley, Neubauer, Marsh....360–9
Nickel Base Indirectly Heated Oxide Cathode—A Review, The. Bounds, Hambleton....102–06
Noise and Crosstalk on Nl Carrier Systems, The Control of. Aikens, Thaeler....605–10; disc. 610
Nonarmored Submarine Telegraph Cable, A. Lawton, Hutchins....153–6; disc. 156
Nondestructive Sensing of Magnetic Cores. Buck, Frank....822–30
Nondestructive Testing of Insulation. Brancato....202–05; disc. 205
Nonlinear Capacitor, A Mathematical Analysis of a Series Circuit Containing a. Pipes....238–42; disc. 242
Nonlinear Electric Circuit Problems, Applications of Integral Equations to the Solution of. Pipes....445–50
Nonsynchronous Tie Between A-C Power Systems, Electronic Frequency Changer Used as. Winograd....263–72; disc. 272
Nonuniform Fields at Radio Frequencies, The Electric Strength of Air in. Whitehead, Birx, Miller..520–4
NTSC Color Television, Transmitting Terminal Apparatus for. Burr....26–33
Nuclear Power Plant Control Considerations. Schultz....22–6
Null-Type Recorder with Balancing Amplifier Which Provides Damping and Suppresses the Quadrature Component, A-C. Williams, Payne....611–15; disc. 615
Numerical Solutions, Temperature Fields in Electrical Coils. Schneider....768–71

O

Oil-Immersed 110-Kv 440-Kw Rectifier Unit for Resnatron Plate Supply. Featherstone, Ling, Cartwright....145–51; disc. 151
Oil, Propagation Mechanism of Impulse Corona and Breakdown in. Liao, Anderson...641–7; disc. 647
Optimum Design of Permanent Magnets Subjected to Demagnetizing Effects, New Method for the. Ziegler....253–62; disc. 262
Oscillation as Well as the Asymmetry of the Resonance Curves, Effects of Harmonics on the Frequency of. Mostafa....309–14
Oscillators, The First-Order Behavior of Separable. dePackh....450–5
Oscillatory Circuits Containing Iron-Cored Inductances. Mostafa, El-Shishini....274–83
Output Formulas for the Resistance Bridge, Sensitivity and. Andress....213–16; disc. 216, 868
Oxide Cathode—A Review, The Nickel Base Indirectly Heated. Bounds, Hambleton....102–06

P

Path Testing for Microwave-Radio Routes. Campbell....326–34
Permeability Analyzer for Magnetic Amplifier Cores, A. Siskind....572–6
Permeability of Silicon-Iron at Very Low Flux Densities, The. Both....656–61; disc. 661
Permanent Magnets Subjected to Demagnetizing Effects, New Method for the Optimum Design of. Ziegler....253–62; disc. 262
Petroleum Industry, Use of Radio and Radar in the. Rust....796–800
Pipe Network Analysis, The Use of Electric Network Analyzers for. Stephenson, Eaton....857–61
Plant Control Considerations, Nuclear Power. Schultz....22–6
Plastic, Forcing Function Generator Employing Conductive. Norman....576–81
Plate Supply, Oil-Immersed 110-Kv 440-Kw Rectifier Unit for Resnatron. Featherstone, Ling, Cartwright....145–51; disc. 151
Pole Lines with Increasing Line Voltage, Some Aspects of Joint Use of Wood. Bloecker, Bullard....709–12
Political Conventions, Television Coverage of the National. Ralston, Wickline....1–14
Polyethylene Terephthalate—Its Use as a Capacitor Dielectric. Wooley, Kohman, McMahon....33–7
Power Meter, 40- to 4,000-Microwatt. Lange....492–4
Power Plant Control Considerations, Nuclear. Schultz....22–6
Power Stations and Substations, Protection of Wire Communication Facilities Serving. Alexander....587–91; disc. 591
Power Systems, Electronic Frequency Changer Used as Nonsynchronous Tie Between A-C. Winograd....263–72; disc. 272
Power Systems with Rectifier Loads, Capacitors in. Schmidt....14–17
Preliminary Development of a Magnettor Current Standard. Felch, Potter....524–31
Principles of Tape-to-Card Conversion in the AMA System. Groth....42–52
Problems by D-C Network Analyzer, The Solution of Sturm-Liouville. Swenson....811–13
Problems on Automatic Computing Equipment, Solutions of Boundary-Value. Verzuh....813–21
Problems to Consider in Applying Selenium Rectifiers. Gramels....488–92
Production of Unipolar Air with Radium Isotopes. Martin....771–6
Program Transmission Over Type-N Carrier Telephone. Case, Kerney....791–5
Progressive Code Digital Quantizer, A. Raasch.567–71
Propagation Mechanism of Impulse Corona and Breakdown in Oil. Liao, Anderson....641–7; disc. 647
Protection of Wire Communication Facilities Serving Power Stations and Substations. Alexander....587–91; disc. 591
Pulse Response Characteristics of Rectangular-Hysteresis-Loop Ferromagnetic Materials. Wylen....648–55; disc. 655

Q

Quadrature Component, A-C Null-Type Recorder with Balancing Amplifier Which Provides Damping and Suppresses the. Williams, Payne....611–15; disc. 615
Quality-Control Measurements, Integrating Instruments for Simplified. Longenecker....375–80
Quantizer, A Progressive Code Digital. Raasch....567–71

R

RCA 3-Gun Shadow-Mask Color Kinescope, The. Law....134–40
Radar in the Petroleum Industry, Use of Radio and. Rust....796–800
(Radio) A Million-Watt Naval Communication Transmitter. Walter....369–74
Radio and Radar in the Petroleum Industry, Use of. Rust....796–800
Radio Communication Systems, A Miniature Compandor for General Use in Wire and. Boxall, Caruthers....804–11
(Radio Communications) Microwave as Applied to Railroad Operation in the Gulf Coast Area. Thomas....800–04

(Radio Communications) Railroad Communications Using High-Power Carrier Technique. Hansen, Dean, Mullins.......785–90; disc. 790
Radio Frequencies, The Electric Strength of Air in Nonuniform Fields at. Whitehead, Birx, Miller.........520–4
(Radio) Frequency Generating Equipment for Million Watt Navy Transmitter. Robertson........475–7
Radio-Frequency Microvoltages, Accurate. Selby.........158–63; disc. 163
Radio Routes, Path Testing for Microwave. Campbell.........326–34
(Radio) The New Jersey Turnpike—A Unique Highway Communication System. Godley, Neubauer, Marsh.........360–9
Radium Isotopes, Production of Unipolar Air with. Martin.........771–6
Railroad Communications Using High-Power Carrier Technique. Hansen, Dean, Mullins.........785–90; disc. 790
Railroad Operation in the Gulf Coast Area, Microwave as Applied to. Thomas.........800–04
Range, Electron-Tube Voltmeter Concepts in the Mid-Frequency. Hoadley.........501–04
Ratio, Computation of Impedance and Efficiency of Transmission Line with High Standing-Wave. Macalpine.........334–9
Rayleigh Method in Network Calculations, The. Schott, Heilfron.........89–93
Reactance on Voltage Regulation of 6-Phase Rectifiers, Influence of A-C. Witzke, Kresser, Dillard.........244–52; disc. 252
Reactors with Inductive D-C Load, Saturable. Storm:
Part II. Transient Response......182–90; disc. 191
Reader, A Subscriber Toll Dialing Tape. Blashfield.........17–21
Recombination at Low Pressures, Electron-Ion. Borowitz.........430–5
Recorded Machine Announcements, Telephone System Applications of. Bennett.........478–83
Recorder with Balancing Amplifier Which Provides Damping and Suppresses the Quadrature Component, A-C Null-Type. Williams, Payne.........611–15; disc. 615
Recording and Accounting in the SATT System, Automatic Call. Ostline.........56–62
Recording, Synchronized Magnetic Tape. Ranger.........581–6
Recording System, An Automatic Transfer Function Measuring and. Ehret, Hochschild, Embree, Grogan.........664–9; disc. 669
Rectangular-Hysteresis-Loop Ferromagnetic Materials, Pulse Response Characteristics of. Wylen.........648–55; disc. 655
Rectangular Hysteresis Loops, Stressed Ferrites Having. Williams, Sherwood, Goertz, Schnettler......531–7
Rectangular Loop Core Materials, An Instability of Self-Saturated Magnetic Amplifiers Using. Batdorf, Johnson.........223–7; disc. 227
Rectifier Installation, Load-Dropping Tests on a Large Ignitron. Pope, Dillard, Marcum.........164–72; disc. 172
Rectifier Loads, Capacitors in Power Systems with. Schmidt.........14–17
Rectifier Unit for Resnatron Plate Supply, Oil-Immersed 110-Kv 440-Kw. Featherstone, Ling, Cartwright.........145–51; disc. 151
Rectifiers, Equivalent Machine Constants for. Dortort.........435–8
Rectifiers, Extended Regulation Curves for 6-Phase Double-Way and Double-Wye. Dortort.........192–8; disc. 198
Rectifiers, Influence of A-C Reactance on Voltage Regulation of 6-Phase. Witzke, Kresser, Dillard.........244–52; disc. 252
Rectifiers, Problems to Consider in Applying Selenium. Gramels.........488–92
Rectifiers, Voltage Regulation of 12-Phase Double-Way. Witzke, Kresser, Dillard.........689–97; disc. 697
Rectifiers, Water Cooling Systems of Mercury-Arc. (Committee Report).........465–75
Regulation Curves for 6-Phase Double-Way and Double-Wye Rectifiers, Extended. Dortort.........192–8; disc. 198
Regulation of 12-Phase Double-Way Rectifiers, Voltage, Witzke, Kresser, Dillard.........689–97; disc. 697
Regulator, Type-N Carrier Telephone Deviation. Perkins, Mahoney.........757–62
Relay Circuits, Automatic Testing of Wired. Hanson.........850–7
Relay Switching Circuits, The Diode Matrix as a Component in. Bush.........833–8
Reliability in Industrial Electronic Equipment. Cook.........351–60
Resistance Bridge, Sensitivity and Output Formulas for the. Andress.........213–16; disc. 216, 868
Resnatron Plate Supply, Oil-Immersed 110-Kv 440-Kw Rectifier Unit for. Featherstone, Ling, Cartwright.........145–51; disc. 151
Resonance Curves, Effects of Harmonics on the Frequency of Oscillation as Well as on the Asymmetry of the. Mostafa.........309–14
Response with Magnetic Amplifiers, Fast. Scorgie.........741–9
Rooms, The Use of Steel Sheet for the Construction of Shielded. Intrator.........599–604; disc. 604

S

SATT System, Automatic Call Recording and Accounting in the. Ostline.........56–62
Saturable Reactors with Inductive D-C Load. Storm:
Part II. Transient Response......182–90; disc. 191
Saturating Magnetic Amplifiers Using Rectangular Loop Core Materials, An Instability of Self-. Batdorf, Johnson.........223–7; disc. 227
Scanning and Memory, An Automatic Control System with Provision for. Young.........392–5
Selector, Type-7 Crossbar. Hutton.........830–3
Selenium Rectifiers, Problems to Consider in Applying. Gramels.........488–92
Sensing of Magnetic Cores, Nondestructive. Buck, Frank.........822–30
Sensitivity and Output Formulas for the Resistance Bridge. Andress.........213–16; disc. 216, 868
Separable Oscillators, The First-Order Behavior of. dePackh.........450–5
Series Circuit Containing a Nonlinear Capacitor, A Mathematical Analysis of a. Pipes.........238–42; disc. 242
Service Standards and Application to Design of a Television Broadcast Network, Development of Television. Reed.........838–50
Shadow-Mask Color Kinescope, The RCA 3-Gun. Law.........134–40
Shaft-Position Transducers, Digital-to-Analogue. O'Neil.........37–41
Sheet for the Construction of Shielded Rooms, The Use of Steel. Intrator.........599–604; disc. 604
Shielded Rooms, The Use of Steel Sheet for the Construction of. Intrator.........599–604; disc. 604
Silicon-Iron at Very Low Flux Densities, The Permeability of. Both.........656–61; disc. 661
Simplified Quality-Control Measurements, Integrating Instruments for. Longenecker.........375–80
Six-Phase Rectifiers, Influence of A-C Reactance on Voltage Regulation of. Witzke, Kresser, Dillard.........244–52; disc. 252
Solution Applied to the Heat Conduction Problem in a Cartridge Fuse, An Analysis of an Analogue. Guile, Carne.........861–7
Solution of Sturm-Liouville Problems by D-C Network Analyzer, The. Swenson.........811–13
Solutions for Vacuum-Tube Circuits, Block-Diagram. Stout.........561–7
Solutions of Boundary-Value Problems on Automatic Computing Equipment. Verzuh.........813–21
Solutions, Temperature Fields in Electrical Coils: Numerical. Schneider.........768–71
Some Aspects of Joint Use of Wood Pole Lines with Increasing Line Voltage. Bloecker, Bullard..709–12
Source Field Problems, A Capacitively Coupled Field Mapper for 2-Dimensional Distributed. Gilbert, Gilbert.........345–9; disc. 349
Square-Loop Core Materials, Theory of Magnetic Amplifiers with. Storm........629–37; disc. 637
Standard, Preliminary Development of a Magnettor Current. Felch, Potter.........524–31
Standards and Application to Design of a Television Broadcast Network, Development of Television Service. Reed.........838–50
Standing-Wave Ratio, Computation of Impedance and Efficiency of Transmission Line with High. Macalpine.........334–9
Stationary Electric Circuits, Basic Concepts in the Analysis of. Spence, Cahn......617–23; disc. 623
Stator Copper Temperature Indicator, Generator. Brownlee, Brown.........676–8; disc. 678
Steel Sheet for the Construction of Shielded Rooms, The Use of. Intrator.........599–604; disc. 604
Strength of Air in Nonuniform Fields at Radio Frequencies, The Electric. Whitehead, Birx, Miller.........520–4
Stressed Ferrites Having Rectangular Hysteresis Loops. Williams, Sherwood, Goertz, Schnettler......531–7
Sturm-Liouville Problems by D-C Network Analyzer, The Solution of. Swenson.........811–13
Submarine Telegraph Cable, A Nonarmored. Lawton, Hutchins.........153–6; disc. 156
Subscriber Toll Dialing Tape Reader, A. Blashfield.........17–21
Sulfur Hexafluoride in Nonuniform Fields, Dielectric Breakdown of. Works, Dakin....682–7; disc. 687
Superregenerative Detectors, A New Method for Treating Electron Tubes when Used as. Mostafa, El-Shishini:
Part I. Superregenerative Circuits Under No-Signal Condition.........207–13
Part II. Superregenerative Circuits Under Signal Conditions.........283–9
Part III. Experimental Investigation of Superregenerative Circuits.........290–7
Switching Circuits, An Application of Boolean Algebra to the Design of Electronic. Washburn......380–8
Switching Circuits, The Diode Matrix as a Component in Relay. Bush.........833–8
Switching System, Transistors and Their Circuits in the 4A Toll Crossbar. Mallery.........388–92
Synchronized Magnetic Tape Recording. Ranger.........581–6
Synchronizer, Cathode-Ray Synchroscope and Automatic. Ahmed.........217–22; disc. 222
Synchros, A Magnetic Amplifier for. Van Allen.........749–56; disc. 756
Synchroscope and Automatic Synchronizer, Cathode-Ray. Ahmed.........217–22; disc. 222
Synthesis of Combinational Logic Circuits, The Map Method for. Karnaugh.........593–8; disc. 598

T

Tachometer, A Commutatorless D-C. Eckels, Peck.........625–8; disc. 628
Tape Reader, A Subscriber Tool Dialing. Blashfield.........17–21
Tape Recording, Synchronized Magnetic. Ranger.........581–6
Tape-to-Card Conversion in the AMA System, Principles of. Groth.........42–52
Technical Characteristics of FTL Type No. 20-B UHF Television Transmitter. Bradburd.........555–61
(Telegraph) A Level Compensator for Telephotograph Systems. Jones, Phelps.........537–41
Telegraph Cable, A Nonarmored Submarine. Lawton, Hutchins.........153–6; disc. 156
Telegraph Industry, Wire and Cable in the. Markley.........298–308; disc. 308
Telephone Carrier Systems, Co-ordination of Ml and Nl. Smith, Cornell, Jerome.........517–20
Telephone Deviation Regulator, Type-N Carrier. Perkins, Mahoney.........757–62
Telephone, Program Transmission Over Type-N Carrier. Case, Kerney.........791–5
Telephone-System Applications of Recorded Machine Announcements. Bennett.........478–83
Telephone Trunks, Transmission Design of Intertoll. Huntley.........670–6; disc. 676
Telephone, Type-ON Carrier. Fracassi, Kahl...713–21
Telephotograph Systems, A Level Compensator for. Jones, Phelps.........537–41
Television Coverage of the National Political Conventions. Ralston, Wickline.........1–14
Television Service Standards and Application to Design of a Television Broadcast Network, Development of. Reed.........838–50
Television Terminals, L3 Coaxial System—. Rieke, Graham.........541–51
(Television) The Measurement of Random Monochrome Video Interference. Barstow, Christopher.........735–41
Television Transmitter, Technical Characteristics of FTL Type No. 20-B UHF. Bradburd......555–61
(Television) Synchronized Magnetic Tape Recording. Ranger.........581–6
(Television) The RCA 3-Gun Shadow-Mask Color Kinescope. Law.........134–40
Television, Transmitting Terminal Apparatus for NTSC Color. Burr.........26–33
Temperature Fields in Electrical Coils: Numerical Solutions. Schneider.........768–71
Temperature Indicator, Generator Stator Copper. Brownlee, Brown.........676–8; disc. 678
Temperatures, Heat Transfer from Electron Tubes at High Altitudes and High Ambient. Buckland.........698–703; disc. 703
Terephthalate—Its Use as a Capacitor Dielectric, Polyethylene. Wooley, Kohman, McMahon..33–7
Terminal Apparatus for NTSC Color Television, Transmitting. Burr.........26–33
Terminals, L3 Coaxial System—Television. Rieke, Graham.........541–51
Testing for Microwave-Radio Routes, Path. Campbell.........326–34
Testing of Insulation, Nondestructive. Brancato.........202–05; disc. 205
Testing of Wired Relay Circuits, Automatic. Hanson.........850–7
Tests on a Large Ignitron Rectifier Installation, Load-Dropping. Pope, Dillard, Marcum.........164–72; disc. 172
Theory of Magnetic Amplifiers with Square-Loop Core Materials. Storm.........629–37; disc. 637
Theory, The Elements of Electrocardiographic. Frank.........125–34
Ticketing System, The XY Toll. Foote.........339–44
Tie Between A-C Power Systems, Electronic Frequency Changer Used as Nonsynchronous. Winograd.........263–72; disc. 272
Titanate Capacitors, The Effect of Minor Constituents in High Dielectric Constant. Coffeen......704–09
Toll Crossbar Switching System, Transistors and Their Circuits in the 4A. Mallery.........388–92

Toll Dialing Tape Reader, A Subscriber. Blashfield........17–21
Toll Messages, Mechanized Billing of AMA. Slade........175–82; disc. 182
Toll Ticketing System, The XY. Foote........339–44
Toll Transmission Equipment, Heat Dissipation from. Coy........762–8
Transducers, Digital-to-Analogue Shaft-Position. O'Neil........37–41
Transfer Function Measuring and Recording System, An Automatic. Ehret, Hochschild, Embree, Grogan........664–9; disc. 669
Transformations in Band-Pass Filters, Maximum Impendance. O'Donnell, Williams........110–13
Transient Analyzer for Magnetic Amplifiers, A. Smith........461–5
Transistors and Their Circuits in the 4A Toll Crossbar Switching System. Mallery........388–92
Transmission Design of Intertoll Telephone Trunks. Huntley........670–6; disc. 676
Transmission Equipment, Heat Dissipation from Toll. Coy........762–8
Tranmission Line with High Standing-Wave Ratio, Computation of Impedance and Efficiency of. Macalpine........334–9
Transmission Over Type-N Carrier Telephone, Program. Case, Kerney........791–5
Transmitter, A Million-Watt Naval Communication. Walter........369–74
Transmitter, Frequency Generating Equipment for Million Watt Navy. Robertson........475–7
Transmitter, Technical Characteristics of FTL Type 20-B UHF Television. Bradburd........555–61
Transmitting Terminal Apparatus for NTSC Color Television. Burr........26–33
Transverse Magnetic Field on an Unconfined Glow Discharge, The Influence of a. McBee, Dow........229–37; disc. 237
Trunks, Transmission Design of Intertoll Telephone. Huntley........670–6; disc. 676
Tube Filament Connections, Gas. Brenner, Schmitt........776–84; disc. 784
Tube Voltmeter Concepts in the Mid-Frequency Range, Electron. Hoadley........501–04
Tubes Having Thoriated-Tungsten Filaments, High-Power Industrial Vacuum. Ayer........121–5
Tungsten Filaments, High-Power Industrial Vacuum Tubes Having Thoriated-. Ayer........121–5
Turnpike—A Unique Highway Communication System, The New Jersey. Godley, Neubauer, Marsh.360–9
Twelve-Phase Double-Way Rectifiers, Voltage Regulation of. Witzke, Kresser, Dillard........689–97; disc. 697
Two-Dimensional Distributed Source Field Problems, A Capactively Coupled Field Mapper for. Gilbert, Gilbert........345–9; disc. 349
Type-N Carrier Telephone Deviation Regulator. Perkins, Mahoney........757–62
Type-N Carrier Telephone, Program Transmission over. Case, Kerney........791–5
Type-ON Carrier Telephone. Fracassi, Kahl...713–21
Type-7 Crossbar Selector. Hutton........830–3

U

UHF Television Transmitter, Technical Characteristics of FTL Type No. 20-B. Bradburd........555–61
Ultrasonic Diathermy, Heating of Fat-Muscle Layers by Electromagnetic and. Schwan, Carstensen, Li........483–8
Unipolar Air with Radium Isotopes, Production of. Martin........771–6
Universal Meter for Measuring Voltages at High Impedances, Micromicroamperes, and Insulation Resistance, A. Clark, Watson, Mergner.....551–5
Use of Electric Network Analyzers for Pipe Network Analysis, The. Stephenson, Eaton........857–61

V

Vacuum-Tube Circuits, Block-Diagram Solutions for. Stout........561–7
Vacuum Tubes Having Thoriated-Tungsten Filaments, High-Power Industrial. Ayer........121–5
Value Problems on Automatic Computing Equipment, Solutions of Boundary-. Verzuh........813–21
Video Interference, The Measurement of Random Monochrome. Barstow, Christopher......735–41
Voltage Regulation of 6-Phase Rectifiers, Influence of A-C Reactance on. Witzke, Kresser, Dillard........244–52; disc. 252
Voltage Regulation of 12-Phase Double-Way Rectifiers. Witzke, Kresser, Dillard ....689–71; disc. 697
Voltages at High Impedances, Micromicroamperes, and Insulation Resistance, A Universal Meter for Measuring. Clark, Watson, Mergner 551–5
Voltmeter Concepts in the Mid-Frequency Range, Electron Tube. Hoadly........501–04

W

Water Cooling Systems of Mercury-Arc Rectifiers. (Committee Report) 465–75
Wattmeters, Correction of Frequency Errors in. Freeman........679–81; disc. 681
Wheatstone Bridge for Admittance Determinations of Highly Conducting Materials at Low Frequencies. Schwan, Sittel........114–21
Wire and Cable in the Telegraph Industry. Markley........298–308; disc. 308
Wire and Radio Communication Systems, A Miniature Compandor for General Use in. Boxall, Caruthers........804–11
Wire Communication Facilities Serving Power Stations and Substations, Protection of. Alexander........587–91; disc. 591
Wired Relay Circuits, Automatic Testing of. Hanson........850–7
Wood Pole Lines with Increasing Line Voltage, Some Aspects of Joint Use of. Bloecker, Bullard........709–12

X

XY Toll Ticketing System, The. Foote........339–4

2. Author Index

Ahmed, M. Abdel-Halim. Cathode-Ray Synchroscope and Automatic Synchronizer. (53-128)........217–22; disc. 223
Aikens, A. J.; C. S. Thaeler. The Control of Noise and Crosstalk on Nl Carrier Systems. (53-226)........605–10; disc. 610
Alexander, T. W., Jr. Protection of Wire Communication Facilities Serving Power Stations and Substations. (53-203)........587–91; disc. 592
Allen, Gordon Y. R. Disc........591
Amey, William G. Disc........615
Anderson, J. G.; T. W. Liao. Propagation Mechanism of Impulse Corona and Breakdown in Oil. (53-280)........641–7; disc. 647
Andress, Philip M. Sensitivity and Output Formulas for the Resistance Bridge. (53-1)........213–16; disc. 216, 868
Armstrong, C. A. Communications for Civil Defense. (53-187)........315–26
Ayer, R. B. High-Power Industrial Vacuum Tubes Having Thoriated-Tungsten Filaments. (53-173)........121–5

B

Barstow, J. M.; H. N. Christopher. The Measurement of Random Monochrome Video Interference (53-356)........735–41
Bashara, N. M. Some Fluorinated Liquid Dielectrics. (53-135)........79–85
Batdorf, S. B.; W. N. Johnson. An Instability of Self-Saturating Magnetic Amplifiers Using Rectangular Loop Core Materials. (53-18) ...223–7; disc. 228
Bennett, W. Telephone-System Applications of Recorded Machine Announcements. (53-219).478–83
Birx, D. L.; J. B. Whitehead, C. F. Miller. The Electric Strength of Air in Nonuniform Fields at Radio Frequencies. (53-262)........520–4
Blashfield, W. H. A Subscriber Toll Dialing Tape Reader. (53-112)........17–21
Bloecker, W. E.; W. R. Bullard. Some Aspects of Joint Use of Wood Pole Lines with Increasing Line Voltage. (53-317)........709–12
Boadway, C. W. Disc........790
Borden, Perry A. Disc........628
Borowitz, Sidney. Electron-Ion Recombination at Low Pressures. (53-263)........430–5
Both, Eberhard. The Permeability of Silicon-Iron at Very Low Flux Densities. (53-259)........656–61; disc. 663
Bounds, A. M.; P. N. Hambleton. The Nickel Base Indirectly Heated Oxide Cathode—A Review. (53-174)........102–06
Boxall, F. S.; R. S. Caruthers. A Miniature Compandor for General Use in Wire and Radio Communication Systems. (53-384)........804–11
Bradburd, E. M. Technical Characteristics of FTL Type No. 20-B UHF Television Transmitter. (53-299)........555–61
Brancato, Emanuel L. Nondestructive Testing of Insulation. (53-125)........202–05; disc. 206
Brenner, Morris; H. G. Schmitt. Gas Tube Filament Connections. (53-357)........776–84
Brewer, E. W. Disc........273
Brown, H. E.; A. L. Brownlee. Generator Stator Copper Temperature Indicator. (53-294)........676–8; disc. 678
Brown, J. S.; W. R. LePage, C. R. Cahn. Analysis of a Comb Filter Using Synchronously Commutated Capacitors. (53-61)........63–8
Brownlee, A. L.; H. E. Brown. Generator Stator Copper Temperature Indicator. (53-294)........676–8; disc. 678
Buck, Dudley, A.; W. I. Frank. Nondestructive Sensing of Magnetic Cores. (53-409)......822–30
Buckland, Mrs. B. O. Heat Transfer from Electron Tubes at High Altitudes and High Ambient Temperatures. (53-307)........698–703; disc. 703
Buckley, Page S. Disc........669
Bullard, W. R.; W. E. Bloecker. Some Aspects of Joint Use of Wood Pole Lines with Increasing Line Voltage. (53-317)........709–12
Burr, Page. Transmitting Terminal Apparatus for NTSC Color Television. (53-91)........26–33
Bush, G. L. The Diode Matrix as a Component in Relay Switching Circuits. (53-403)........833–8

C

Cahn, C. R.; W. R. LePage, J. S. Brown. Analysis of a Comb Filter Using Synchronously Commutated Capacitors. (53-61)........63–8
Cahn, C. R.; D. W. Spence. Basic Concepts in the Analysis of Stationary Electric Circuits. (53-268)........617–23; disc. 624
Caldwell, S. H. Disc........598
Calvert, J. F. Disc........350
Camilli, G.; R. E. Plump. Fluorine-Containing Gaseous Dielectrics. (53-126)........93–102
Camilli, G.; R. E. Plump, T. W. Liao. Disc.......687
Campbell, R. D. Path Testing for Microwave-Radio Routes. (53-191)........326–34
Carstensen, Edwin L.; H. P. Schwan. Application of Electric and Acoustic Impedance Measuring Techniques to Problems in Diathermy. (53-137)........106–10
Carne, E. B.; A. E. Guile. An Analysis of an Analogue Solution Applied to the Heat Conduction Problem in a Cartridge Fuse. (53-411)........861–7
Carstensen, Edwin L.; H. P. Schwan, K. Li. Heating of Fat Muscle Layers by Electromagnetic and Ultrasonic Diathermy. (53-206)........483–8
Cartwright, P. A.; R. P. Featherstone, C. Ling. Oil-Immersed 110-Kv 440-Kw Rectifier Unit for Resnatron Plate Supply. (53-67)....145–51; disc. 152
Caruthers, R. S.; F. S. Boxall. A Miniature Compandor for General Use in Wire and Radio Communication Systems. (53-384)........804–11
Case, R. L.; I. Kerney. Program Transmission Over Type-N Carrier Telephone. (53-335)........791–5
Chegwidden, R. A. Disc........662
Christopher, H. N.; J. M. Barstow. The Measurement of Random Monochrome Video Interference. (53-356)........735–41
Cioffi, P. P. Disc........262
Clark, W. R.; R. E. Watson, G. C. Mergner. A Universal Meter for Measuring Voltages at High Impedances, Micromicroamperes, and Insulation Resistance. (53-298)........551–5
Coffeen, W. W. The Effect of Minor Constituents in High Dielectric Constant Titanate Capacitors. (53-236)........704–09
Coggeshall, I. S. Disc........157
Cook, Ellsworth D. Reliability in Industrial Electronic Equipment. (53-208)........351–60
Cornell, L. P., Jr.; E. P. Smith, M. G. Jerome. Coordination of M1 and N1 Telephone Carrier Systems. (53-225)........517–20
Coy, J. A. Heat Dissipation from Toll Transmission Equipment. (53-344)........762–8
Crawshaw, R. C. Disc........676

D

Dakin, T. W. Disc........206
Dakin, T. W.; C. N. Works. Dielectric Breakdown of Sulful Hexafluoride in Nonuniform Fields (53-235)........682–7; disc. 687
Davidson, J. L.; R. E. Koll. A New Design 60-Cycle A-C Network Analyzer. (53-160)........141–5
Dean, W. W. H.; A. K. Hansen, F. G. Mullins, Jr. Railroad Communications Using High-Power Carrier Technique. (53-342)....785–90; disc. 791
dePackh, David C. The First-Order Behavior of Separable Oscillators. (53-267)........450–5
Dickinson, F. R.; L. H. Morris, G. H. Lovell. L3 Coxial System—Amplifiers. (53-222)......505–17
Dieterly, D. C. Disc........227, 661
Dillard, J. K.; S. J. Pope, C. R. Marcum. Load-Dropping Tests on a Large Ignitron Rectifier Installation. (53-65)........164–72; disc. 174

Dillard, J. K.; R. L. Witzke, J. V. Kresser. Influence of A-C Reactance on Voltage Regulation of 6-Phase Rectifiers. (53-54)........244–52; disc. 252
Dillard, J. K.; R. L. Witzke, J. V. Kresser. Voltage Regulation of 12-Phase Double-Way Rectifiers. (53-242)........................689–97; disc. 697
Dillard, J. K.; R. L. Witzke, J. V. Kresser. Disc...198
Dortort, I. K. Equivalent Machine Constants for Rectifiers. (53-271)..............................435–8
Dortort, I. K. Extended Regulation Curves for 6-Phase Double-Way and Double-Wye Rectifiers. (53-36) ....192–8; disc. 202
Dortort, I. K. Disc.....................252, 273, 697
Dow, W. G.; W. D. McBee. The influence of a Transverse Magnetic Field on an Unconfined Glow Discharge. (53-154)...............229–37; disc. 238
Duna, A. L. Disc.............................308

E

Early, A. C.; H. L. Smith. Disc..................237
Eaton, J. R.; R. E. Stephenson. The Use of Electric Network Analyzers for Pipe Network Analysis. (53-410)...............................857–61
Eck, Robert N.; P. M. Fischer. Disc...............784
Eckels, A. R.; W. R. Peck. A Commutatorless D-C Tachometer. (53-292)...........625–8; disc. 628
Eckert, E. R. G. Disc..........................703
Ehrbar, R. D.; C. H. Elmendorf, R. H. Klie, A. J. Grossman. The L3 Coaxial System. (53-221).. 395–413
Ehret, R. J.; E. F. Hochschild, J. M. Embree, E. C. Grogan. An Automatic Transfer Function Measuring and Recording System. (53-293)........ ....664–9; disc. 669
Elmendorf, C. H.; R. D. Ehrbar, R. H. Klie, A. J. Grossman. The L3 Coaxial System. (53-221).. 395–413
El-Shishini, M.; Abd El-Samie Mostafa. A New Method for Treating Electron Tubes when Used as Superregenerative Detectors:
  Part I. Superregenerative Circuits Under No-Signal Condition. (53-158)....................207–13
  Part II. Superregenerative Circuits Under Signal Conditions. (53-157)......................283–9
  Part III. Experimental Investigation of Superregenerative Circuits. (53-155)............290–7
El-Shishini, M.; A. El-Samie Mostafa. Oscillatory Circuit Containing Iron-Cored Inductances. (53-156)...............................274–83
Embree, J. M.; R. J. Ehret, E. F. Hochschild, E. C. Grogan. An Automatic Transfer Function Measuring and Recording System. (53-293)........ ....664–9; disc. 669
Estoppey, R. F. Disc...........................681

F

Farrell, W. C. Disc............................610
Featherstone, R. P.; C. Ling, P. A. Cartwright. Oil-Immersed 110-Kv 440-Kw Rectifier Unit for Resnatron Plate Supply. (53-67)....145–51; disc. 152
Felch, E. P.; J. L. Potter. Preliminary Development of a Magnettor Current Standard. (53-291).. 524–31
Finch, T. R.; R. W. Ketchledge. L3 Coaxial System—Equalization and Regulation. (53-223)....414–30
Finzi, L. A.; R. A. Mathias. An Application of Magnetic Amplifier Circuits to Perform Multiplication and Other Analytical Operations. (53-285).455–61
Fischer, Paul M.; R. N. Eck. Disc................784
Fooks, J. H. Disc.............................350
Foord, T. R. Disc.............................687
Foote, H. L. The XY Toll Ticketing System. (53-184) 339–44
Fracassi, R. D.; H. Kahl. Type-ON Carrier Telephone. (53-333)..........................713–21
Frank, Ernest. The Elements of Electrocardiographic Theory. (53-136).......................125–34
Frank, Werner I.; D. A. Buck. Nondestructive Sensing of Magnetic Cores. (53-409)..............822–30
Freeman, James Robert. Correction of Frequency Errors in Wattmeters. (53-295)..679–81; disc. 682

G

Gilbert, Edward O.; E. G. Gilbert. A Capacitively Coupled Field Mapper for 2-Dimensional Distributed Source Field Problems. (53-193)...... ....345–9; disc. 351
Gilbert, Elmer G.; E. O. Gilbert. A Capacitively Coupled Field Mapper for 2-Dimensional Distributed Source Field Problems. (53-193)...... ....345–9; disc. 351
Gilbert, J. J. Disc.............................156
Godley, P. F.; J. R. Neubauer, D. R. Marsh. The New Jersey Turnpike—A Unique Highway Communication System. (53-213)..............360–9
Goertz, Matilda; H. J. Williams, R. C. Sherwood, F. J. Schnettler. Stressed Ferrites Having Rectangular Hysteresis Loops. (53-260).........531–7
Graham, R. S.; J. W. Rieke. L3 Coaxial System—Television Terminals. (53-224)............541–51
Gramels, J. Problems to Consider in Applying Selenium Rectifiers. (53-215)...................488–92
Grogan, E. C.; R. J. Ehret, E. F. Hochschild, J. M. Embree. An Automatic Transfer Function Measuring and Recording System. (53-293)......... ....664–9; disc. 669
Grossman, A. J.; C. H. Elmendorf, R. D. Ehrbar, R. H. Klie. The L3 Coaxial System. (53-221)..395–413
Groth, Willard B. Principles of Tape-to-Card Conversion in the AMA System. (53-113).......42–52
Gulle, A. E.; E. B. Carne. An Analysis of an Analogue Solution Applied to the Heat Conduction Problem in a Cartridge Fuse. (53-411)..............861–7
Gutzwiller, W. E. Disc.........................272

H

Hale, J. W.; F. R. Richardson. Mathematical Descriptions of Core Losses. (53-261).......495–501
Halloran, J. J. Disc............................152
Halperin, Herman. Disc........................678
Hambleton, P. N.; A. M. Bounds. The Nickel Base Indirectly Heated Oxide Cathode—A Review. (53-174)...............................102–06
Hansen, A. K.; W. W. H. Dean, F. G. Mullins, Jr. Railroad Communications Using High-Power Carrier Technique. (53-342)....785–90; disc. 791
Hanson, A. N. Automatic Testing of Wired Relay Circuits. (53-407)........................850–7
Harris, F. K.; F. Wenner. Disc...................216
Hawthorne, Edward I. Flow of Energy in D-C Machines. (53-266)...........................438–45
Heilfron, Jack; F. W. Schott. The Rayleigh Method in Network Calculations. (53-62)...........89–93
Herskind, C. C. Disc...........................172
Higgins, Thomas J. Disc.......................615
Hill, Donald M. Disc..........................163
Hoadley, George B. Electron Tube Voltmeter Concepts in the Mid-Frequency Range. (53-308).. 501–04
Hoadley, George B. Disc.......................616
Hochschild, E. F.; R. J. Ehret, J. M. Embree, E. C. Grogan. An Automatic Transfer Function Measuring and Recording System. (53-293)........ ....664–9; disc. 669
Horsch, J. R.; G. W. Penney, E. A. Sack. Dielectric Amplifiers. (53-130)......................68–79
House, C. B. Flux Preset High-Speed Magnetic Amplifiers. (53-337)........................728–35
Huntley, H. R. Transmission Design of Intertoll Telephone Trunks. (53-227).....670–6; disc. 676
Hutchins, L. H., Jr.; C. S. Lawton. A Nonarmored Submarine Telegraph Cable. (53-163).......... ....153–6; disc. 157
Hutton, Robert W. Type-7 Crossbar Selector. (53-408)...................................830–3

I

Intrator, A. M. The Use of Steel Sheet for the Construction of Shielded Rooms. (53-197)....599–604

J

Jeanne, P. A. Disc............................272
Jerome, M. G.; E. P. Smith, L. P. Cornell, Jr. Coordination of M1 and N1 Telephone Carrier Systems. (53-225).......................517–20
Johnson, L. J.; S. E. Rauch. Disc................756
Johnson, W. N.; S. B. Batdorf. An Instability of Self-Saturating Magnetic Amplifiers Using Rectangular Loop Core Materials. (53-18)....223–7; disc. 228
Johnston, R. F.; A. G. Ratz. A Graphical Method for Flip-Flop Design. (53-60)..................52–6
Jones, T. A.; W. A. Phelps. A Level Compensator for Telephotographic Systems. (53-297).......537–41

K

Kahl, H.; R. D. Fracassi. Type-ON Carrier Telephone. (53-333)..........................713–21
Kang, B. P. Disc.............................206
Karnaugh, M. The Map Method for Synthesis of Combinational Logic Circuits. (53-217)........ ....593–8; disc. 599
Kassimir, William J. Disc......................222
Kerney, Iden; R. L. Case. Program Transmission Over Type-N Carrier Telephone. (53-335)..791–5
Ketchledge, R. W.; T. R. Finch. L3 Coaxial System—Equalization and Regulation. (53-223)....414–30
Klie, R. H.; C. H. Elmendorf, R. D. Ehrbar, A. J. Grossman. The L3 Coaxial System. (53-221).. 395–413
Kohman, G. T.; M. C. Wooley, W. McMahon. Polyethylene Terephthalate—Its Use as a Capacitor Dielectric. (53-6).........................33–7
Koll, Robert E.; J. L. Davidson. A New Design 60-cycle A-C Network Analyzer. (53-160).....141–5
Kresser, J. V.; R. L. Witzke, J. K. Dillard. Influence of A-C Reactance on Voltage Regulation of 6-Phase Rectifiers. (53-54)........244–52; disc. 252
Kresser, J. V.; R. L. Witzke, J. K. Dillard. Voltage Regulation of 12-Phase Double-Way Rectifiers. (53-242)......................689–97; disc. 697
Kresser, J. V.; R. L. Witzke, J. K. Dillard. Disc...198
Ku, Y. H. Disc..............................243

L

Lange, R. W. 40- to 4,000-Microwatt Power Meter. (53-281).................................492–4
Law, H. B. The RCA 3-Gun Shadow-Mask Color Kinescope. (53-175)......................134–40
Lawton, C. S.; L. H. Hutchins, Jr. A Nonarmored Submarine Telegraph Cable. (53-163)......... ....153–6; disc. 157
Legg, V. E. Disc.............................662
LePage, W. R.; C. R. Cahn, J. S. Brown. Analysis of a Comb Filter Using Synchronously Commutated Capacitors. (53-61)........................63–8
Li, Kam, H. P. Schwan, E. L. Carstensen. Heating of Fat Muscle Layers by Electromagnetic and Ultrasonic Diathermy. (53-206)................483–8
Liao, T. W.; J. G. Anderson. Propagation Mechanism of Impulse Corona and Breakdown in Oil. (53-280)..........................641–7; disc. 647
Liao, T. W.; G. Camilli, R. E. Plump. Disc........687
Ling, C.; R. P. Featherstone, P. A. Cartwright. Oil-Immersed 110-Kv 440-Kw Rectifier Unit for Resnatron Plate Supply. (53-67)............ ....145–51; disc. 152
Longenecker, D. M. Integrating Instruments for Simplified Quality-Control Measurements. (53-205)...................................375–80
Lord, Harold W. Dynamic Hysteresis Loops of Several Core Materials Employed in Magnetic Amplifiers. (53-117)..................................85–8
Lord, Harold W. The Influence of Magnetic Amplifier Circuitry Upon the Operating Hysteresis Loops. (53-311)..................................721–8
Lord, Harold W. Disc..........................637
Lovell, G. H.; L. H. Morris, F. R. Dickinson. L3 Coaxial System—Amplifiers. (53-222).....505–17

M

Macalpine, W. W. Computation of Impedance and Efficiency of Transmission Line with High Standing-Wave Ratio. (53-192)....................334–9
Mack, D. R. Disc............................623
Mahoney, J. J.; E. H. Perkins. Type-N Carrier Telephone Deviation Regulator. (53-334)..757–62
Mallery, P. Transistors and Their Circuits in the 4A Toll Crossbar Switching System. (53-218)..388–92
Marcum, C. R.; S. J. Pope, J. K. Dillard. Load-Dropping Tests on a Large Ignitron Rectifier Installation. (53-65)...........164–72; disc. 174
Markley, W. F. Wire and Cable in the Telegraph Industry. (53-164)............298–308; disc. 308
Marsh, D. R.; P. F. Godley, J. R. Neubauer. The New Jersey Turnpike—A Unique Highway Communication System. (53-213)..............360–9
Martin, Thomas L., Jr. Production of Unipolar Air with Radium Isotopes. (53-351)............771–6
Mathias, R. A.; L. A. Finzi. An Application of Magnetic Amplifier Circuits to Perform Multiplication and Other Analytical Operations. (53-285)..455–61
McBee, W. D.; W. G. Dow. The Influence of a Transverse Magnetic Field on an Unconfined Glow Discharge. (53-154)...........229–37; disc. 238
McMahon, W.; M. C. Wooley, G. T. Kohman. Polyethylene Terephthalate—Its Use as a Capacitor Dielectric. (53-6).........................33–7
Meahl, Harry M. Disc..........................604
Mergner, G. C.; W. R. Clark, R. E. Watson. A Universal Meter for Measuring Voltages at High Impedances, Micromicroamperes, and Insulation Resistance. (53-298).......................551–5
Miller, C. F.; J. B. Whitehead, D. L. Birx. The Electric Strength of Air in Non-uniform Fields at Radio Frequencies. (53-262)....................520–4
Moore, A. D. Disc............................350
Morris, L. H.; G. H. Lovell, F. R. Dickinson. L3 Coaxial System—Amplifiers. (53-222).....505–17
Morton, Lysle W. Disc.........................173
Mostafa, Abd El-Samie. Effects of Harmonics on the Frequency of Oscillation as Well as on the Asymmetry of the Resonance Curves. (53-159)...309–14

Mostafa, Abd El-Samie; M. El-Shishini. A New Method for Treating Electron Tubes when Used as Superregenerative Detectors:
  Part I. Superregenerative Circuits Under No-Signal Condition. (53-158)......................207–13
  Part II. Superregenerative Circuits Under Signal Conditions. (53-157)......................283–9
  Part III. Experimental Investigation of Superregenerative Circuit. (53-155)..............290–7
Mostafa, Abd El-Samie; M. El-Shishini. Oscillatory Circuits Containing Iron-Cored Inductances. (53-156).................................274–83
Mullins, F. G., Jr.; A. K. Hansen, W. W. H. Dean. Railroad Communications Using High-Power Carrier Technique. (53-342)....785–90; disc. 791

N

Neubauer, J. R.; P. F. Godley, D. R. Marsh. The New Jersey Turnpike—A Unique Highway Communication System. (53-213)..............360–9
Norman, L. W. Forcing Function Generator Employing Conductive Plastic. (53-289)..........576–81

O

O'Donnell, T. J.; E. M. Williams. Maximum Impedance Transformations in Band-Pass Filters. (53-143)................................110–13
Oliver, F. S. Disc.................................205
O'Neil, S. J. Digital-to-Analogue Shaft-Position Transducers. (53-7).......................37–41
Ostline, J. E. Automatic Call Recording and Accounting in the SATT System. (53-111)..........56–62

P

Payne, J. F., Jr.; A. J. Williams, Jr. A-C Null-Type Recorder with Balancing Amplifier Which Provides Damping and Suppresses the Quadrature Component. (53-244)..............611–15; disc. 616
Peck, W. R.; A. R. Eckels. A Commutatorless D-C Tachometer. (53-292)..........625–8; disc. 628
Penney, G. W.; J. R. Horsch, E. A. Sack. Dielectric Amplifiers. (53-130)......................68–79
Perkins, E. H.; J. J. Mahoney. Type-N Carrier Telephone Deviation Regulator. (53-334)..757–62
Phelps, W. A.; T. A. Jones. A Level Compensator for Telephotograph Systems. (53-297)........537–41
Phillips, R. A. Disc.........................191, 639
Pipes, Louis A. A Mathematical Analysis of a Series Circuit Containing a Nonlinear Capacitor. (53-29)......................238–42; disc. 243
Pipes, L. A. Applications of Integral Equations to the Solution of Nonlinear Electric Circuit Problems. (53-265)................................445–50
Plump, R. E.; G. Camilli. Fluorine-Containing Gaseous Dielectrics. (53-126)................92–102
Plump, R. E.; G. Camilli, T. W. Liao. Disc.......687
Pope, S. J.; J. K. Dillard, C. R. Marcum. Load-Dropping Tests on a Large Ignitron Rectifier Installation. (53-65)...........164–72; disc. 174
Potter, J. L.; E. P. Felch. Preliminary Development of a Magnettor Current Standard. (53-291)..524–31
Potter, W. F. Disc.................................610
Pritchett, Wilson. Disc.............................682

R

Raasch, Floyd. A Progressive Code Digital Quantizer. (53-287)................................567–71
Ralston, R. W.; B. D. Wickline. Television Coverage of the National Political Conventions. (53-90)..1–14
Ranger, R. H. Synchronized Magnetic Tape Recording. (53-300)............................581–6
Ratz, A. G.; R. F. Johnston. A Graphical Method for Flip-Flop Design. (53-60)..................52–6
Rauch, S. E.; L. J. Johnson. Disc...............756
Reed, O. W. B., Jr. Development of Television Service Standards and Application to Design of a Television Broadcast Network. (53-414).......838–50
Riaz, M. Disc.................................191
Richardson, F. R.; J. W. Hale. Mathematical Descriptions of Core Losses. (53-261)......495–501
Rieke, John W.; R. S. Graham. L3 Coaxial System—Television Terminals. (53-224)..........541–51
Robertson, D. G. Frequency Generating Equipment for Million Watt Navy Transmitter. (53-198)..475–7
Rogers, Walter E. Disc.........................349
Rust, W. M., Jr. Use of Radio and Radar in the Petroleum Industry. (53-381)...........796–800

S

Sack, E. A.; G. W. Penney, J. R. Horsch. Dielectric Amplifiers. (53-130)......................68–79
Schlawin, R. G. Disc..........................151
Schmidt, A., Jr. Capacitors in Power Systems with Rectifier Loads. (53-24)..................14–17
Schmitt, H. G.; M. Brenner. Gas Tube Filament Connections. (53-357)..................776–84
Schneider, P. J. Temperature Fields in Electrical Coils: Numerical Solutions. (53-352).....768–71
Schnettler, F. J.; H. J. Williams, R. C. Sherwood, M. Goertz. Stressed Ferrites Having Rectangular Hysteresis Loops. (53-260)...............531–7
Schonland, B. F. J. Disc........................647
Schott, Frederick W.; J. Heilfron. The Rayleigh Method in Network Calculations. (53-62).. 89–93
Schultz, M. A. Nuclear Power Plant Control Considerations. (53-31).........................22–6
Schwan, Herman P.; E. L. Carstensen. Application of Electric and Acoustic Impedance Measuring Techniques to Problems in Diathermy. (53-137)...106–10
Schwan, Herman P.; E. L. Carstensen, K. Li. Heating of Fat Muscle Layers by Electromagnetic and Ultrasonic Diathermy. (53-206)...........483–8
Schwan, Herman P.; K. Sittel. Wheatstone Bridge for Admittance Determinations of Highly Conducting Materials at Low Frequencies. (53-127)...114–21
Scorgie, D. G. Fast Response with Magnetic Amplifiers. (53-336)..........................741–9
Scorgie, D. G. Disc............................637
Seeley, H. T. Disc............................592
Selby, Myron C. Accurate Radio-Frequency Microvoltages. (53-22)..............158–63; disc. 164
Sherwood, R. C.; H. J. Williams, M. Goertz, F. J. Schnettler. Stressed Ferrites Having Rectangular Hysteresis Loops. (53-260)...............531–7
Simmons, W. R. Disc...........................350
Siskind, Philip. A Permeability Analyzer for Magnetic Amplifier Cores. (53-290)...............572–6
Sittel, Karl; H. P. Schwan. Wheatstone Bridge for Admittance Determinations of Highly Conducting Materials at Low Frequencies. (53–127)...114–21
Slade, Frank D. Mechanized Billing of AMA Toll Messages. (53-114)............175–82; disc. 182
Smith, Arthur Bessey. Disc......................182
Smith, E. J. A Transient Analyzer for Magnetic Amplifiers. (53-283)..........................461–5
Smith, E. P.; L. P. Cornell, Jr., M. G. Jerome. Coordination of M1 and N1 Telephone Carrier Systems. (53-225)..........................517–20
Smith, H. L.; A. C. Early. Disc.................237
Smith, J. J. Disc..............................623
Sommerman, G. M. L. Disc.....................647
Spence, D. W.; C. R. Cahn. Basic Concepts in the Analysis of Stationary Electric Circuits. (53-268)..........................617–23; disc. 624
Stephenson, Robert E.; J. R. Eaton. The Use of Electric Network Analyzers for Pipe Network Analysis. (53-410)................................857–61
Storm, H. F. Saturable Reactors with Inductive D-C Load: ....
Part II. Transient Response. (53-20)..182–90; disc. 192
Storm, H. F. Theory of Magnetic Amplifiers with Square-Loop Core Materials. (53-284)..................................629–37; disc. 640
Stout, T. M. Block Diagram Solutions for Vacuum Tube Circuits. (53-269)...................561–7
Stout, Thomas M. Disc.........................242
Swenson, George W., Jr. The Solution of Sturm-Liouville Problems by D-C Network Analyzer. (53-412)................................811–13

T

Thaeler, C. S.; A. J. Aikens. The Control of Noise and Crosstalk on N1 Carrier Systems. (53-226).................................605–10; disc. 610
Thomas, L. R. Microwave as Applied to Railroad Operation in the Gulf Coast Area. (53-382)...800–04

U

Uchrin, George C. Disc.........................262

V

Van Allen, R. L. A Magnetic Amplifier for Synchros. (53-339)......................749–56; disc. 756
Verzuh, Frank M. Solutions of Boundary-Value Problems on Automatic Computing Equipment. (53-395)................................813–21

W

Walter, J. C. A Million-Watt Naval Communication Transmitter. (53-199)...................369–74
Ward, C. E. Disc..............................655
Washburn, S. H. An Application of Boolean Algebra to the Design of Electronic Switching Circuits. (53-216)................................380–8
Watson, R. E.; W. R. Clark, G. C. Mergner. A Universal Meter for Measuring Voltages at High Impedances, Micromicroamperes, and Insulation Resistance. (53-298)......................551–5
Wenner, F.; F. K. Harris. Disc.................216
Whitehead, J. B.; D. L. Birx, C. F. Miller. The Electric Strength of Air in Nonuniform Fields at Radio Frequencies. (53-262)...................520–4
Wickline, B. D.; R. W. Ralston. Television Coverage of the National Political Conventions. (53-90)...1–14
Williams, A. J., Jr.; J. F. Payne, Jr. A-C Null-Type Recorder with Balancing Amplifier Which Provides Damping and Suppresses the Quadrature Component. (53-244).............611–15; disc. 616
Williams, E. M.; T. J. O'Donnell. Maximum Impedance Transformations in Band-Pass Filters. (53-143)................................110–13
Williams, H. J.; R. C. Sherwood, M. Goertz, F. J. Schnettler. Stressed Ferrites Having Rectangular Hysteresis Loops. (53-260)...............531–7
Winograd, Harold. Electronic Frequency Changer Used as Nonsynchronous Tie Between A-C Power Systems. (53-53)..............263–72; disc. 273
Winograd, H. Disc.............................174
Witzke, R. L.; J. V. Kresser, J. K. Dillard. Influence of A-C Reactance on Voltage Regulation of 6-Phase Rectifiers. (53-54)............244–52; disc. 252
Witzke, R. L.; J. V. Kresser, J. K. Dillard. Voltage Regulation of 12-Phase Double-Way Rectifiers. (53-242)......................689–97; disc. 697
Witzke, R. L.; J. V. Kresser, J. K. Dillard. Disc...198
Wooley, M. C.; G. T. Kohman, W. McMahon. Polyethylene Terephthalate—Its Use as a Capacitor Dielectric. (53-6)..........................33–7
Works, C. N.; T. W. Dakin. Dielectric Breakdown of Sulfur Hexafluoride in Nonuniform Fields. (53-235)..........................682–7; disc. 687
Wyatt, Kenneth S. Disc........................157
Wylen, Joseph. Pulse Response Characteristics of Rectangular-Hysteresis-Loop Ferromagnetic Materials. (53-286)..............648–55; disc. 655

Y

Young, N. H. An Automatic Control System with Provision for Scanning and Memory. (53-202)..392–5

Z

Ziegler, Hans. K. New Method for the Optimum Design of Permanent Magnets Subjected to Demagnetizing Effects. (53-152)...253–62; disc. 262

Preface

The AIEE *Transactions* for 1953 (volume 72) is published in three parts: Part I. Communication and Electronics; Part II. Applications and Industry; and Part III. Power Apparatus and Systems. The papers in each of the three parts are classified according to subject matter as follows:

Part I. Communication and Electronics

Communication Switching Systems
Radio Communications Systems
Special Communications Applications
Telegraph Systems
Television and Aural Broadcasting Systems
Wire Communications Systems
Basic Sciences
Computing Devices
Electrical Techniques in Medicine and Biology
Electronic Power Converters
Electronics
Instruments and Measurements
Magnetic Amplifiers
Metallic Rectifiers
Nucleonics

Part II. Applications and Industry

Air Transportation
Domestic and Commercial Applications
Land Transportation
Marine Transportation
Production and Application of Light
Chemical, Electrochemical and Electrothermal Applications
Electric Heating
Electric Welding
Feedback Control Systems
General Industry Applications
Industrial Control
Industrial Power Systems
Mining and Metal Industry

Part III. Power Apparatus and Systems

Carrier Current
Insulated Conductors
Power Generation
Protective Devices
Relays
Rotating Machinery
Substations
Switchgear
System Engineering
Transformers
Transmission and Distribution

Education
Safety
Research

Each part has been indexed separately in the back of that particular part. The three parts are not cross-referenced; hence the user should determine first whether the subject matter of the paper desired is in the field of communication and electronics, applications and industry, or power apparatus and systems. Papers are listed in the subject index under several key words in the titles. The original numbers assigned to the papers are given in the author index.

Volume 72 contains the technical papers and related discussions presented at these meetings:

1. Winter General Meeting, New York, N. Y., January 19–23, 1953.
2. Southern District Meeting, Louisville, Ky., April 22–24, 1953.
3. North Eastern District Meeting, Boston, Mass., April 29–May 1, 1953.
4. Summer General Meeting, Atlantic City, N. J., June 15–19, 1953.
5. Pacific General Meeting, Vancouver, B. C., Canada, September 1–4, 1953.
6. Middle Eastern District Meeting, Charleston, W. Va., September 29–October 1, 1953.
7. Aircraft Electric Equipment Conference, Seattle, Wash., September 30–October 2, 1953.
8. Fall General Meeting, Kansas City, Mo., November 2–6, 1953.

Statements and opinions given in the papers and discussions published in *Transactions* are the expressions of the contributors for which the American Institute of Electrical Engineers assumes no responsibility.

Television Coverage of the National Political Conventions

R. W. RALSTON
NONMEMBER AIEE

B. D. WICKLINE
NONMEMBER AIEE

IF ONE were to indicate in a single word the over-all characteristic of the 1952 Democratic and Republic political conventions, it would be difficult to avoid choosing "quantity." From a political standpoint, certainly at no other event in recent history was manifest the number of candidates, the number of delegates, the number of committees, or the number of commentaries by a not inconsiderable group of reporters. However, the most striking application of this word is to the means by which the first large-scale television coverage of these events was provided over such an extended period of time. The concentration of equipment and personnel in Chicago, Ill., during this period was unprecedented for any event, making the television camera as all-pervading as was the microphone at previous conventions. This was most graphically illustrated at the Republican convention after General Eisenhower was selected as candidate. There were cameras at his suite in the Blackstone Hotel, cameras to cover his exit from the hotel as he departed on his trip to the Amphitheatre, cameras along the street as he made his journey, cameras outside the Amphitheatre, ad infinitum.

The provision by Illinois Bell Telephone Company of television facilities for this period represented no exception to the application of this word. As an operating company in the Bell System, it was the company's job to provide all television channels connecting various pickup points, such as the International Amphitheatre, the Conrad Hilton Hotel, and the Blackstone Hotel, to the network terminus in Chicago, and also to provide virtually all other local television channels which were used for indirect connection to the networks. Naturally, this unusually severe requirement introduced many problems not previously encountered. Arranging for the installation and simultaneous operation of 11 microwave equipments at one location, increasing the operating technical forces by a factor of over two—these are but a few examples.

Preliminary plans for the political conventions were formulated immediately after the International Amphitheatre had been officially selected as the site, at which time telecasters indicated to our Commercial Department that two connecting channels would be required. However, these plans were not crystallized until a more detailed résumé of video channel requirements was received early in February 1952. This indicated that approximately 25 channels additional to those in use in the Chicago area at that time would be required during the month of July. Included were 19 channels between the International Amphitheatre—the primary source of convention activities—and the following locations:

1. The Merchandise Mart, location of the studios of television station *WNBQ*, the Chicago National Broadcasting Company (NBC) outlet.
2. The Civic Opera Building, location of the studios and transmitter of television station *WENR-TV*, the Chicago American Broadcasting Company (ABC) outlet; also the transmitter location of station *WNBQ*.
3. The Hilton Hotel, a secondary point of convention coverage.
4. Number 2 Toll Office, the terminating point in the Loop area for video transmission to or reception from the cross-country network.

The remaining six channels extended between various miscellaneous locations in the Loop area.

This was an unprecedented request for service; it would double the number of channels then being provided by our company. However, this requirement was understandable in view of the broadcasters' planning. Each contemplated providing convention coverage for the entire day during the progress of the conventions. This would consist not only of programs originating at the Amphitheatre, but also from the Hilton Hotel, various convention headquarters, and from the network. Consequently each planned to establish an independent master control point at the Amphitheatre through which most of these programs would pass regardless of whether they originated at the Amphitheatre. Editing and switching of pictures would take place here prior to transmission for direct telecast to all network stations providing convention coverage. As is evident, this resulted in a relatively enormous request for channels.

A comprehensive review was then made by the Transmission Engineering Division of our company to determine the most suitable means of furnishing the required service. In general terms, of the 19 channels originating or terminating at the Amphitheatre, it was planned to use special video conductors, together with associated video amplifiers for eight of the channels and to use microwave facilities for the remainder. In regard to other locations, conductors would be used where installed or where a reasonably economic rearrangement would make them available. Microwave channels would be used for the remainder. These plans were then discussed with various organizations within the company to implement the required work.

Cable

Before discussing the problem of providing special video pairs, a brief exposition will be made of our cable facilities for television service. Generally speaking, two types of conductors (pairs) are used with amplifying equipment for providing television service. The first is the polyethylene shielded video (PSV) pair, a special video pair. It consists of two 16-gauge conductors, surrounded by polyethylene string and tape insulation and a double shielding of helically wound copper tape. This make-up provides excellent characteristics for television application such as relatively small loss, very low noise level, and excellent shielding. The PSV pair is placed within a protective sheath along with other conductors of its type and regular telephone pairs, thus constituting what is called video cable.

The second type is the ordinary telephone pair. This has from three to five times the loss of the PSV pair, thus requiring more amplification per channel and providing relatively high susceptibility to interference from other telephone pairs. Statistically speaking, paper pairs have relatively high ambient noise when evaluated according to television requirements. Use of these pairs is rapidly being discontinued as video cable is installed.

Paper 53-90, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted October 23, 1952; made available for printing December 17, 1952.

R. W. Ralston and B. D. Wickline are with the Illinois Bell Telephone Company, Chicago, Ill.

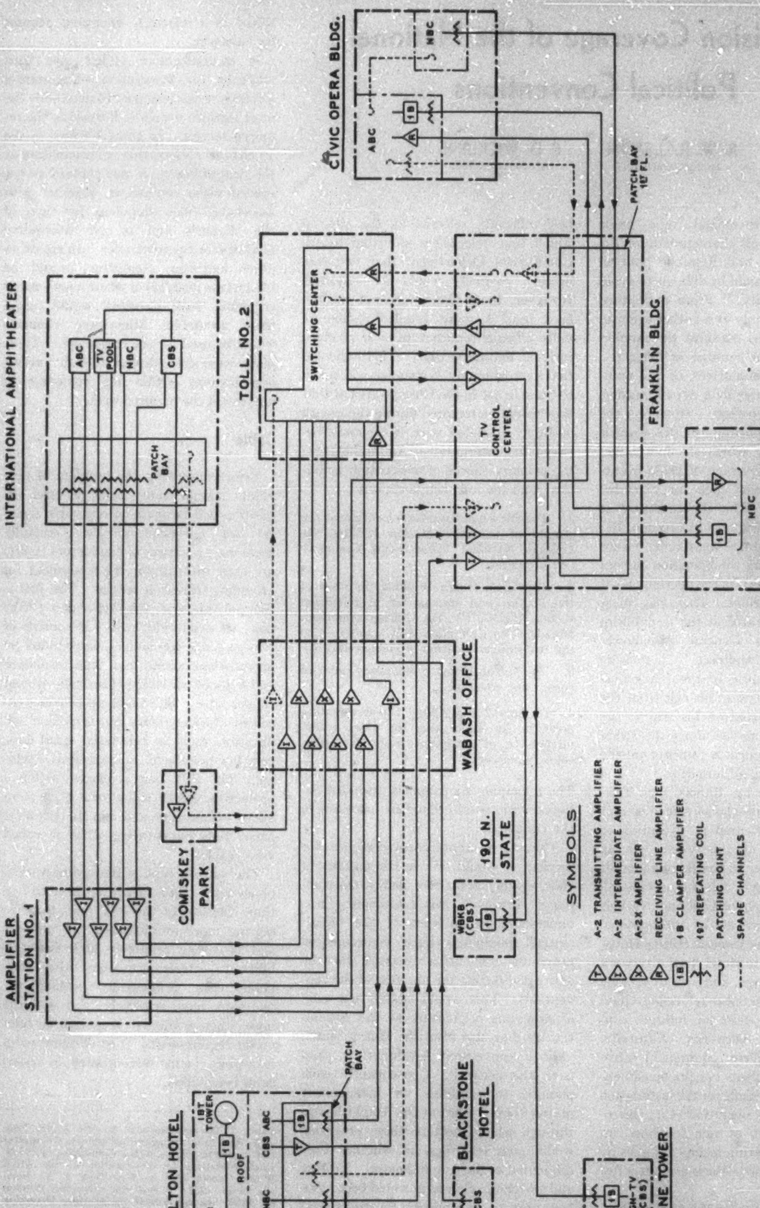

INTERNATIONAL AMPHITHEATER
ABC
TV POOL
NBC
CBS
PATCH BAY
TOLL NO. 2
SWITCHING CENTER
CIVIC OPERA BLDG.
ABC
1B
NBC
PATCH BAY 1st FL.
FRANKLIN BLDG
TV CONTROL CENTER
MERCHANDISE MART
NBC
1B
AMPLIFIER STATION NO. 1
COMISKEY PARK
WABASH OFFICE
190 N. STATE
WBKB (CBS)
1B
SYMBOLS
A-2 TRANSMITTING AMPLIFIER
A-2 INTERMEDIATE AMPLIFIER
A-2X AMPLIFIER
RECEIVING LINE AMPLIFIER
1B CLAMPER AMPLIFIER
197 REPEATING COIL
PATCHING POINT
SPARE CHANNELS
HILTON HOTEL
NBC TOWER
IBT TOWER
1B
ROOF
ABC NBC
CBS ABC
1B
PATCH BAY
BLACKSTONE HOTEL
CBS
TRIBUNE TOWER
1B
WGN-TV (CBS)

Figure 1. Wire channel layout for 1952 political conventions

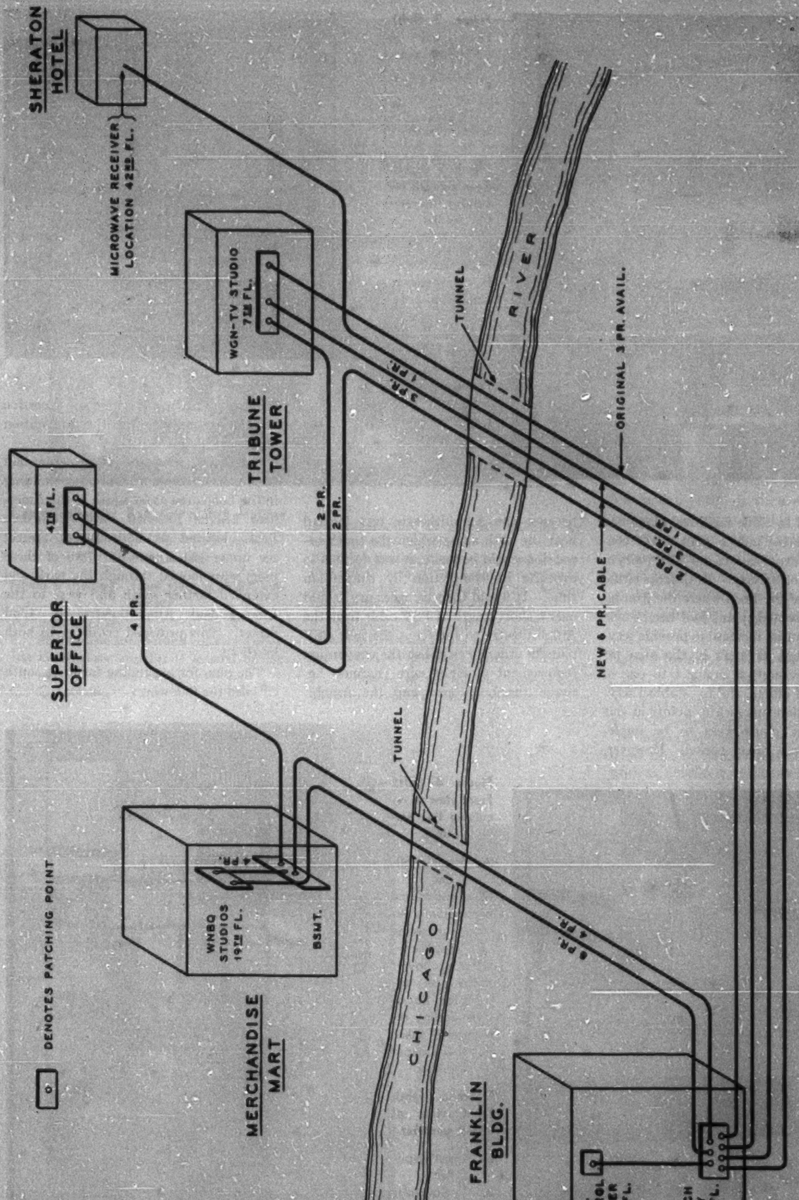

SHERATON HOTEL
MICROWAVE RECEIVER LOCATION 42ND FL.
WGN-TV STUDIO 7TH FL.
TRIBUNE TOWER
4TH FL.
SUPERIOR OFFICE
2 PR.
2 PR.
4 PR.
1 PR.
3 PR.
TUNNEL
RIVER
ORIGINAL 3 PR. AVAIL.
NEW 6 PR. CABLE
1 PR.
3 PR.
2 PR.
TUNNEL
CHICAGO
DENOTES PATCHING POINT
MERCHANDISE MART
WNBQ STUDIOS 19TH FL.
BSMT.
4 PR.
6 PR.
4 PR.
FRANKLIN BLDG.
T.V. CONTROL CENTER 18TH FL.
PATCH BAY 1ST FL.

Figure 2. Layout of cable serving the Merchandise Mart, Superior Office, Tribune Tower, and Sheraton Hotel. Cable entering the Franklin Building from other locations is not shown

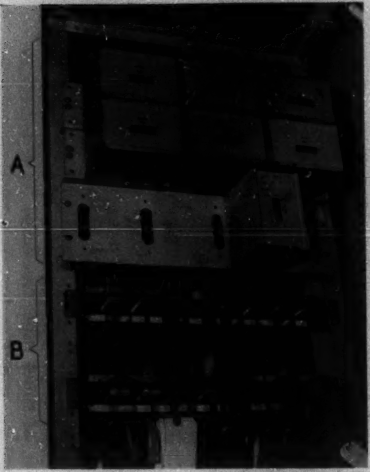


Figure 3 (left). Patch bay at Hilton Hotel

A. 197 coil installation

B. Plugs and jacks for patching

Figure 5 (right). Rear view of receiving line amplifier

A. Input amplifier

B. Impedance conversion device consisting of amplifier—cathode follower shown on extreme right. Equalizers occupy remainder of space

C. Output amplifier

In regard to cable facilities, plans had been formulated independent of the convention for expansion of our PSV pairs in many areas of the city. In the case of the area south of the Loop where the Amphitheatre is located, plans had been made for a substantial increase to provide service to various theaters in the area for theater television, to connect to one of our new toll offices, and to establish service to miscellaneous pickup points in this area. These plans were to be implemented over a broad period. However, because the Amphitheatre was located near the path along which the south expansion would be made, it was decided to expedite its completion by convention time. It would then be necessary to add only a short length of PSV cable from the Amphitheatre to intercept the new PSV pairs in order to establish the convention requirement of eight wire channels between the Loop area and the Amphitheatre. This would avoid an excessive capital expenditure for the convention alone.

Originally, a 4-pair PSV cable had been placed from a central office building in the Loop area as far south as the American League Baseball Park (Comiskey Park) located at 35th Street South; see upper left Figure 1. Two of these pairs were looped through the park and extended farther south and west to the International Amphitheatre at 42nd Street. This provided coverage to both locations.

The plan for expanding facilities south entailed the following:

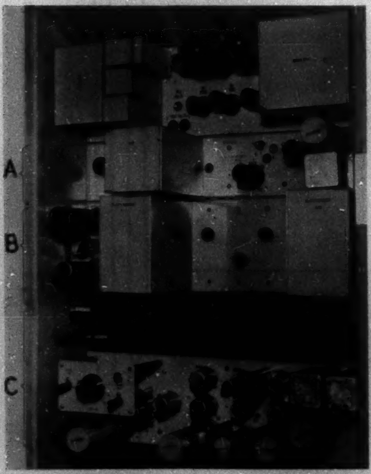


Figure 4 (left). Front view of receiving line amplifier

A. Input amplifier

B. Impedance conversion device consisting of amplifier—cathode follower shown on extreme left. Equalizers occupy remainder of space

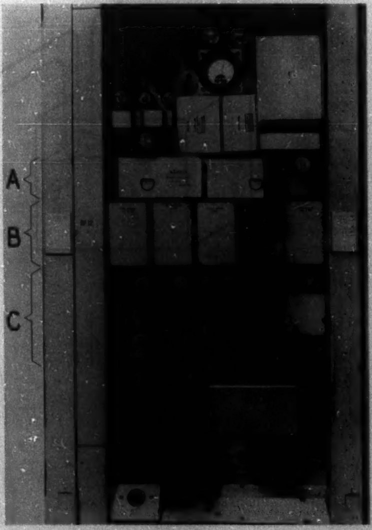
C. 1-B clamper amplifier

Figure 6 (right). Front view of A-2 amplifier

A. Input amplifier

B. Equalizer panel

C. Output amplifier



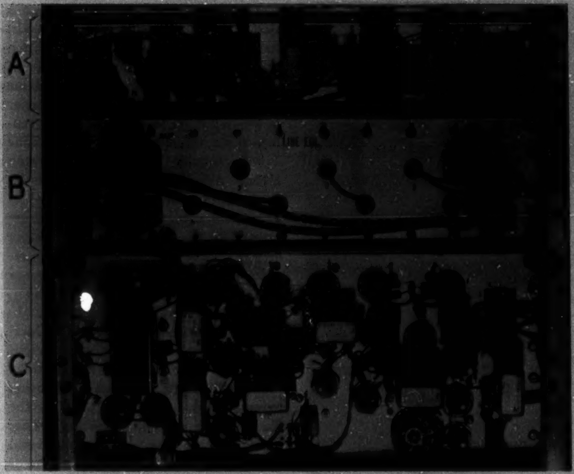


Figure 7 (above). Rear view of A-2 amplifier

A. Input amplifier B. Equalizer panel
C. Output amplifier

Figure 8 (below). Front view of A-2X amplifier

A. Input amplifier
B. Intermediate amplifier and equalizer panel. The intermediate amplifier is located on the upper center
C. Output amplifier

1. Construction of a small building for housing video amplifiers, to be known as Amplifier Station number 1, located about five miles south of the Loop on the fundamental video wire route. The Amphitheatre lies just a few blocks west.

2. Placement of a 4-pair PSV cable south of Wabash Office directly to the new Amplifier Station, by-passing Comiskey Park.

3. Extension of two of the original four pairs to Amplifier Station number 1, by-passing Comiskey Park. This then provided a total of six pairs to the Amplifier Station.

4. Placement of an additional 6-pair extension from the station to the Amphitheatre.

This provided the necessary eight pairs from the loop area to the Amphitheatre, two of which looped through Comiskey Park and did not extend through the Amplifier Station.

In regard to the Hilton Hotel, a rearrangement and addition to existing facilities was sufficient to provide five pairs into the Hilton Hotel, four of which were video pairs and one paper; see Figure 1.

Just north of the loop area it was necessary to reinforce facilities to serve the studios of *WNBQ* at the Merchandise Mart. The method of providing this addition illustrates some fundamental concepts employed in designing our Chicago Video Cable Network and therefore we feel is of interest. Many pickup points served by our cable pairs are used only a few times a year. In these cases, it is our policy to have flexible cable arrangements permitting some pairs to serve more than one location. Similarly, where a number of pairs serve a location in consistent use but where there are occasionally more severe demands for service, additional

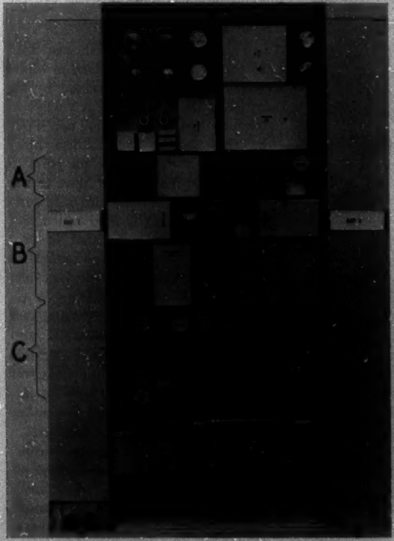


Figure 9 (below). Rear view of A-2X amplifier

A. Input amplifier
B. Intermediate amplifier and equalizer panel. The intermediate amplifier is located on the upper center. Strip on rear may be used for patching equalizers or intermediate amplifiers as required
C. Output amplifier

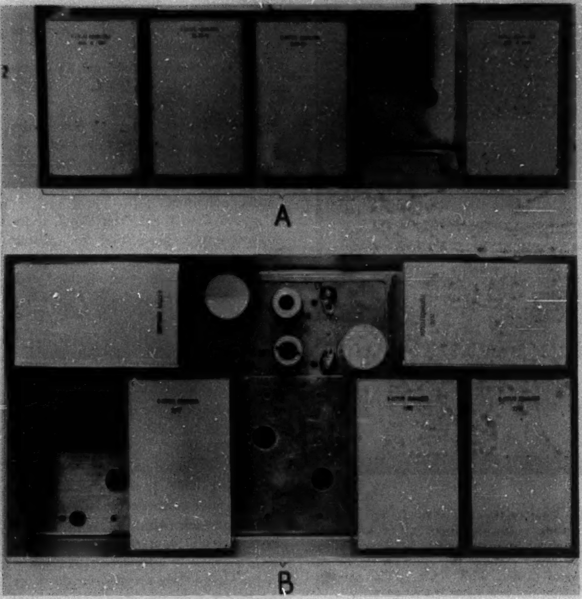


Figure 10 (above). Comparison of original A-2 amplifier and section modified to convert it to an A-2X amplifier

A. Equalizer panel for A-2
B. Equalizer panel and intermediate amplifier for A-2X

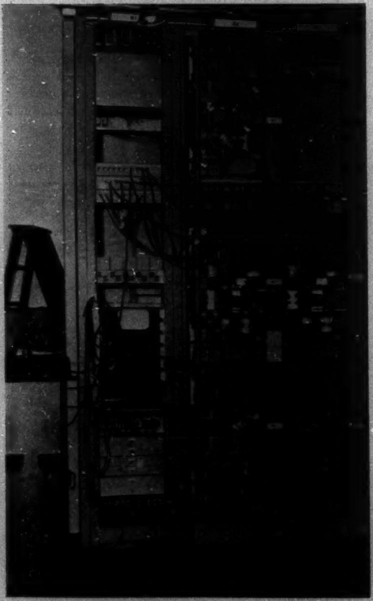


Figure 11 (left). Installation of A-2X amplifiers at Wabash Central Office. On lower left is shown means for monitoring, patching, and communication

pairs are installed, where possible, on a shared basis with other locations. In consequence of this policy, the addition to Mart facilities involved a number of locations in its vicinity: the Tribune Tower where the studios and transmitter of *WGN-TV* are located, one of our central office buildings near the Tribune Tower (Superior Office), and the Sheraton Hotel. As in the case of the Amphitheatre, a plan for additional facilities had been considered previously but was expedited for the conventions since certain of the locations mentioned would require additional pairs in another year or two.

Originally eight PSV pairs were established to the studios of *WNBQ* at the Merchandise Mart. However, these eight pairs were a part of a 12-pair PSV cable, four of which extended past this location to serve the studio of *WGN-TV*, the Sheraton Hotel, and Superior Office; see Figure 2. The plan for additional facilities included the following:

1. Four PSV pairs previously by-passing the Merchandise Mart were looped through the basement at this location and terminated in a patching arrangement, permitting these pairs to be either routed to the *WNBQ* Studios on the 19th floor or looped through to serve the original locations.

2. A new 6-pair PSV cable was established along another path to serve *WGN-TV* and other locations in its vicinity. Three of these pairs were actually terminated at *WGN-TV*, making a total of 8 pairs, while the remaining three were used to facilitate rearrangements at Superior Office and the Sheraton Hotel.

This permitted simultaneous usage of 12 pairs to the Merchandise Mart, in which case *WGN-TV* would be served by six pairs along the new path. A change in patching would permit routing two *WGN* pairs through the Merchandise Mart basement, thus providing eight pairs to *WGN-TV* and limiting the Mart to the use of ten pairs. This also provided flexibility, thus fulfilling a need which had been felt for some time, that of providing alternate routing for channels in the event of a failure of a particular cable.

At the Civic Opera Building, a tie cable consisting of four PSV pairs was placed between the *WENR-TV* and *WNBQ* locations to permit more flexible use of facilities extending from the Civic Opera Building in the event of an unexpectedly high requirement for either telecaster, see Figure 1.

A rearrangement which was accomplished at one of our central office buildings in the Loop requires some background material. The centralized nature of the Loop Area in Chicago, which contains all of the studios and transmitters of television stations, has permitted us to establish one location—the Franklin Building—through which virtually all television channels are routed, see Figure 2. Prior to the rearrangement to be discussed, all channels were routed through a location on the 16th floor called the Television Control Center (TCC). This center is equipped with special video amplifiers, a patch bay and monitoring equipment which permits control of all channels.

However, before the conventions, a new type of amplifying equipment, the Receiving Line Amplifier, to be discussed in the next section, was made available for use on certain short video channels. This piece of equipment obviated the need of many of the video amplifiers, designated type *A-2*, located at the TCC. Since the capabilities of these *A-2* units were not exploited on short links and also since many additional *A-2*'s were needed during the conventions, it was desirable to replace them with Receiving Line Amplifiers. However, it was found that the losses incurred on some of these short channels merely looping in and out of the TCC (1st floor to 16th floor and back) were sufficient to preclude the use of Receiving Line Amplifiers. For this reason arrangements were made to bring all cables entering the Franklin Building to a patching bay on the first floor, thus permitting channels to either enter or bypass the TCC. It was planned to retain control on channels for monitoring by bridging across the line through a high-impedance amplifier at the first floor to the TCC when required. The reason for citing this rearrangement is that virtually all television channels then in existence were affected by this plan, requiring intricate time-consuming co-ordination work for completion.

Cable rearrangements were also made on television channels extending between the Franklin Building and the number 2 Toll Building to permit more flexible arrangement of pairs between all locations. See Figure 1.

Equipment

Wire

To clarify some of the equipment problems involved, some characteristics of wire video channels will be discussed.

A wire video channel consists of conductors, terminal equipment to permit connection to the broadcasters equipment, suitable amplifying equipment, and a means of equalizing the conductor losses which vary with frequency. At the

Figure 12. Telephone equipment room at International Amphitheatre. To the left behind cabinet is patch bay showing jacks and plugs. Above this is shown rear of 197 coil installation. 1-B clamper amplifiers appear on top and bottom of bay respectively. Portable monitor and oscilloscope are shown on dolly. On table are some of the 1-B clamper amplifiers. At right is shown video monitor

Figure 13. Microwave space on roof of International Amphitheatre. R-f cabinet removed from dish second from right

Figure 14. Front view of microwave equipment on roof of the International Amphitheatre

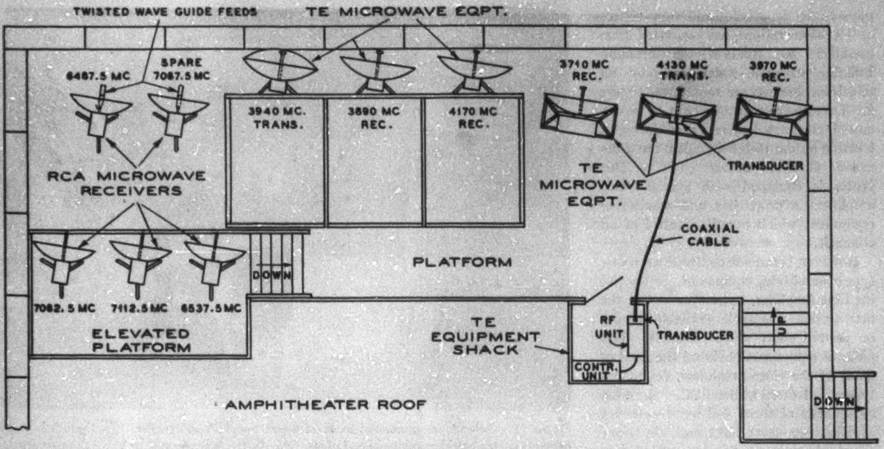

TWISTED WAVE GUIDE FEEDS
TE MICROWAVE EQPT.
6487.5 MC
SPARE 7087.5 MC
3940 MC. TRANS.
3890 MC REC.
4170 MC REC.
3710 MC REC.
4130 MC TRANS.
3970 MC REC.
RCA MICROWAVE RECEIVERS
TE MICROWAVE EQPT.
TRANSDUCER
COAXIAL CABLE
DOWN
PLATFORM
7082.5 MC
7112.5 MC
6537.5 MC
ELEVATED PLATFORM
TE EQUIPMENT SHACK
RF UNIT
TRANSDUCER
CONTR. UNIT
UP
DOWN
AMPHITHEATER ROOF

Figure 15. Block diagram of microwave equipment on roof of the International Amphitheatre

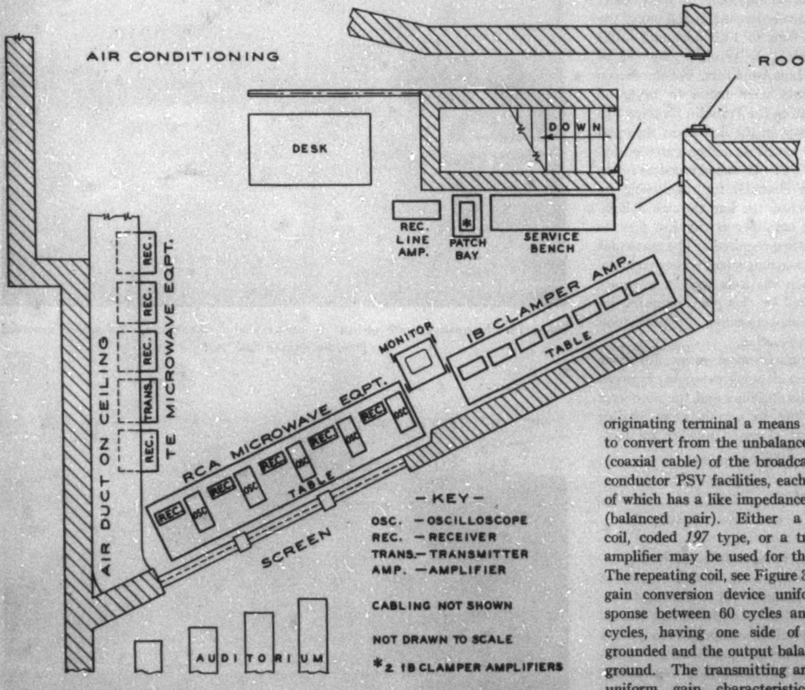

AIR CONDITIONING
ROOF
DESK
DOWN
REC. LINE AMP.
PATCH BAY
SERVICE BENCH
REC.
REC.
REC.
TRANS.
REC.
TE MICROWAVE EQPT.
AIR DUCT ON CEILING
1B CLAMPER AMP.
TABLE
MONITOR
RCA MICROWAVE EQPT.
REC.
OSC
REC.
OSC
REC.
OSC
REC.
OSC
REC.
OSC
REC.
OSC
TABLE
SCREEN
-KEY-
OSC. —OSCILLOSCOPE
REC. —RECEIVER
TRANS.—TRANSMITTER
AMP. —AMPLIFIER
CABLING NOT SHOWN
NOT DRAWN TO SCALE
*2 1B CLAMPER AMPLIFIERS
AUDITORIUM

Figure 16. Block diagram of telephone equipment room in the International Amphitheatre

originating terminal a means is required to convert from the unbalanced facilities (coaxial cable) of the broadcasters to 2-conductor PSV facilities, each conductor of which has a like impedance to ground (balanced pair). Either a repeating coil, coded *197* type, or a transmitting amplifier may be used for this purpose. The repeating coil, see Figure 3, is a 1 to 1 gain conversion device uniform in response between 60 cycles and 7 megacycles, having one side of the input grounded and the output balanced above ground. The transmitting amplifier has uniform gain characteristics between 3 cycles and 4.5 megacycles. It is used

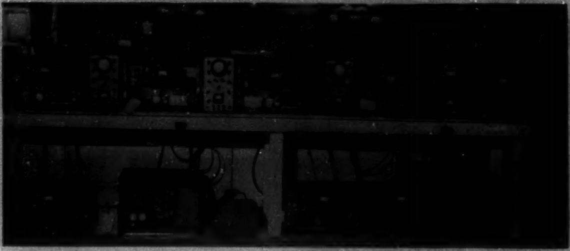


Figure 17. RCA microwave equipment control units and associated oscilloscopes located in telephone equipment room at the International Amphitheatre

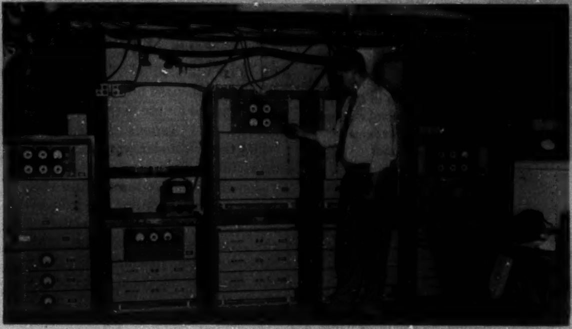


Figure 18. TE control equipment located in telephone equipment room at the International Amphitheatre

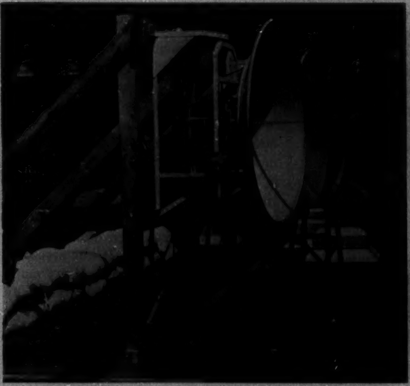
whenever gain is desired, either to increase low signal input or to override high ambient noise such as might exist on paper pairs.

The terminal equipment at the receiving location consists of the repeating coil just mentioned and a *1-B* clamper amplifier; see Figures 4 and 5. The latter consists of an amplifier having a gain of 18 decibels flat to 7 megacycles and a clamper unit which maintains the voltage level of the horizontal synchronizing pulses at a fixed d-c value. This d-c reinsertion is necessary in view of the transformer characteristics of the *197* coil.

Amplification and equalization are accomplished using *A-2* amplifiers; see Figures 6 and 7. Each consists of three components: an input amplifier, an equalizer panel, and an output amplifier. Both input and output may be adapted for connection to coaxial conductors or balanced conductors. In addition, these components may be used in different combinations depending upon the particular need. If it is desired to use the *A-2* system for transmitting purposes where an unbalanced input and a balanced output are required, the input amplifier and equalizer panel are omitted. When the equalizer and the input amplifier are included, the *A-2* is called an intermediate amplifier. The input and output amplifier may make up both gain and equalization on cable losses up to 65 decibels at 4 megacycles. In the case of PSV pairs, this would permit an interval of 4 miles between adjacent amplifiers. To summarize, a wire line channel normally consists of the *197* coil at the originating end, a cable pair, intermediate amplifier, and a coil and clamper at the receiving point.

Figure 19 (lower left). TE microwave equipment located at the Merchandise Mart showing mounting details

Figure 20 (lower right). Microwave Control equipment located at the Merchandise Mart. Cables at right lead out to platform



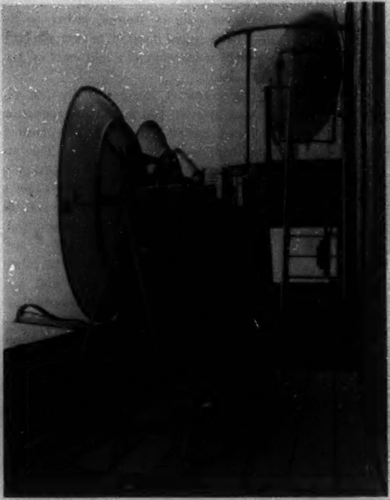


Figure 21 (left). TE microwave equipment located on 44th floor parapet at the Civic Opera Building. Cupola at upper right contains WNBQ microwave equipment

Figure 22 (above). RCA microwave equipment located on top of the Hilton Hotel. Control equipment is located underneath this roof

To provide amplification and equalization on the new wire lines extending south of the Loop area for the convention period, a new item of equipment was made available, the *A-2X* amplifier; see Figures 8 and 9. This had gain and equalization potentials in excess of the *A-2* amplifier, thus permitting adjacent amplifiers to be spaced $5\frac{1}{2}$ miles apart instead of 4 miles. The advantage of this item is readily apparent in view of the relatively inexpensive modifications of the *A-2* amplifiers necessary to convert it to an *A-2X* amplifier. Figure 10 shows the portion modified. The additional gain and equalization are obtained by adding an additional amplifier to the *A-2* system between the input and output amplifiers and providing additional space for equalizers on the equalizer panel. Thus the *A-2X* provides gain and equalization on circuits up to 85 decibels of 4-megacycle loss as compared to 65 decibels for the *A-2* system. Installations of these amplifiers were made at the Wabash Office in the Loop area and at Amplifier Station Number 1. This station was constructed at a location based upon the spacing which could be tolerated between adjacent *A-2X* amplifiers. The Wabash installation is shown in Figure 11.

Another item of equipment which was installed at a number of locations was the Receiving Line Amplifier, Figures 4 and 5. This item, which is compact, and most important, more economical than the *A-2* amplifier, was designed for use at the receiving terminal of relatively short channels. Since it provides both equalization and gain it obviates the need for the *A-2* amplifier on channels of 35 decibels of 4-megacycle loss or less. It is made up of a modified input amplifier from an *A-2* system, an equalizer panel, a special impedance conversion device, and a *1-B* clamper. The impedance conversion device is necessary to change the equalizer panel impedance to that of the *1-B* clamper.

Because of the critical need for *A-2* amplifiers during the convention period, installations of these Receiving Line Amplifiers were expedited at the Civic Opera Building, the Franklin Building, and Number 2 Toll Office. Several portable units were also ordered for use where necessary at miscellaneous locations.

Installations of *197* repeating coils and *1-B* clampers were made at the Hilton Hotel, International Amphitheatres, and other locations. Portable clampers were ordered for use at miscellaneous locations. See Figures 3 and 12.

MICROWAVE

To provide the required number of microwave channels it was necessary to obtain microwave equipment from other operating companies to supplement our own supply. Accordingly, four telephone equipment (TE) type microwave equipments were rented from the Wisconsin Telephone Company and two Radio Corporation of America (RCA) units were rented from the New York Telephone Company. The TE-type equipment operates at 4,000 megacycles and the RCA operates at approximately 7,000 megacycles. Together with the equipment owned or ordered by our company, this provided a total of six TE and nine RCA type equipments. Eleven of these units were planned for use between the Amphitheatre and various Loop locations, while the remainder were to be utilized as maintenance spares and for the provision of regular and unforeseen requirements occurring during the convention period. Various test and maintenance equipment such as oscillators, signal generators, and measuring devices were also obtained for the operation of both wire and microwave channels.

Space

As explained previously, it is company policy to have most channels extended through the Franklin Building for control and monitoring purposes. As applied to a typical microwave pickup, this means that the microwave transmitter would be located at or near the pickup point while the receiver would be located at one of our tall central office buildings in the Loop area. From this point the signal would be transmitted to the broadcaster by a wire channel through the Franklin Building. This is done to enable us to operate the microwave equipment at our locations as much as possible. However, because of the unusual nature of the demand for channels during the convention, it was necessary to establish microwave circuits directly between the pickup point and the broadcaster's location since sufficient wire lines were not available. To implement this plan it was necessary to secure space and roof rights at many more locations than would normally be necessary. The following locations were involved:

1. The International Amphitheatre (all microwave and wire channels).

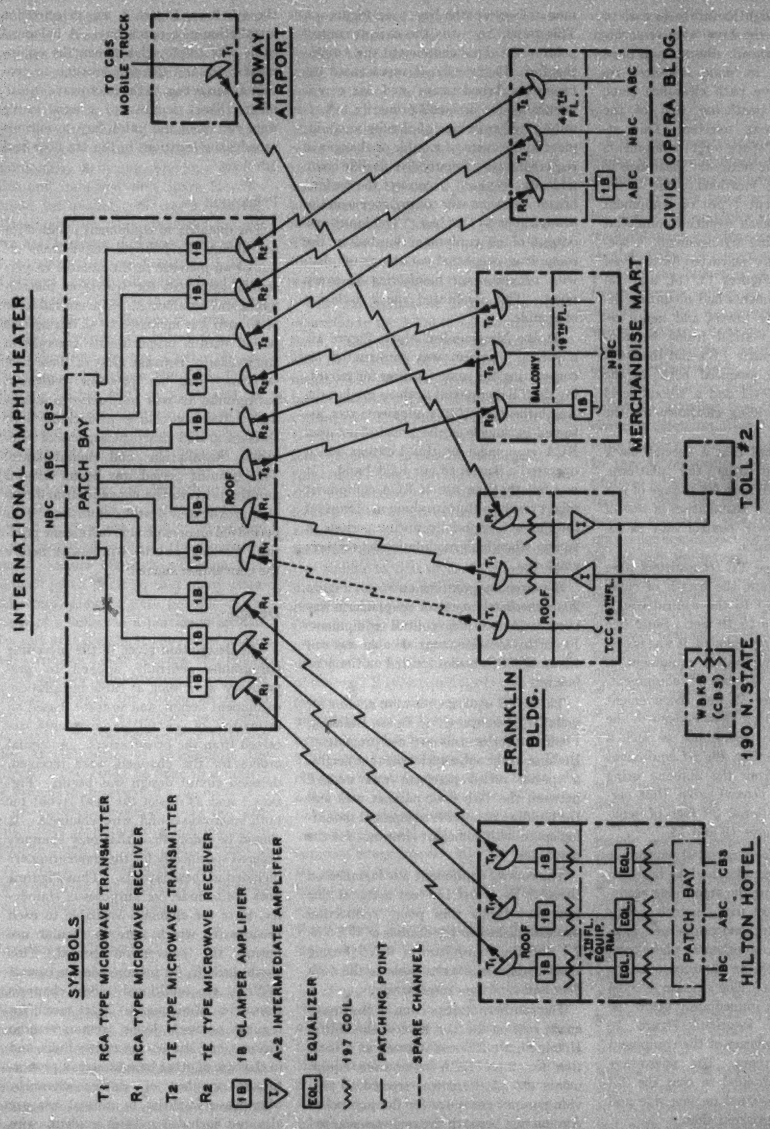

INTERNATIONAL AMPHITHEATER
NBC
ABC
CBS
PATCH BAY
ROOF
1B
R_1
R_2
T_2
TO CBS MOBILE TRUCK
T_1
MIDWAY AIRPORT
CIVIC OPERA BLDG.
44TH FL.
ABC
NBC
ABC
MERCHANDISE MART
BALCONY
19TH FL.
NBC
FRANKLIN BLDG.
ROOF
16TH FL.
TCC
TOLL #2
WBKB (CBS)
190 N. STATE
HILTON HOTEL
ROOF
4TH FL. EQUIP. RM.
EQL.
PATCH BAY
NBC
ABC
CBS
SYMBOLS
T_1 RCA TYPE MICROWAVE TRANSMITTER
R_1 RCA TYPE MICROWAVE RECEIVER
T_2 TE TYPE MICROWAVE TRANSMITTER
R_2 TE TYPE MICROWAVE RECEIVER
1B 1B CLAMPER AMPLIFIER
A-2 INTERMEDIATE AMPLIFIER
EQL. EQUALIZER
197 COIL
PATCHING POINT
SPARE CHANNEL

Figure 23. Block diagram of microwave equipment layout

2. The Civic Opera Building (ABC and NBC microwave channels).

3. Merchandise Mart (NBC microwave channels).

4. Hilton Hotel (microwave and wire channels).

The International Amphitheatre, where 11 microwave channels and eight wire channels would originate and terminate, imposed the most unusual requirements. Here it was necessary to secure space for 11 microwave equipments on the roof, space for associated control equipment within a distance of 200 feet of the roof, and finally space for amplifiers and other equipment for the wire channels. Surveys were conducted to determine what location on the Amphitheatre roof would provide line of sight coverage to each of the locations in the Loop area requiring a microwave channel. Based upon these surveys, which in some instances required microwave path clearance tests, the northwest penthouse roof of the Amphitheatre was selected. This afforded sufficient space for 11 equipments and provided the necessary clearance to each of the loop locations. Due to the fact that different types of equipment would be used, each requiring somewhat different mounting arrangements, a special platform was constructed for this roof location. See Figures 13, 14, and 15. Three TE equipments had mounting details for portable towers and hence required that a vertical frame be constructed. The other TE's and the RCA equipment were mounted both on the main platform level and a second level above this, providing clearance for the parabolas.

It will be noted that a special shack was constructed atop the platform. This was intended both for storage of test equipment and for installation of one of the TE microwave control and radio-frequency (r-f) units.

The earlier type TE transmitters were so constructed that the length of cable connecting the r-f to the control equipment was limited to 10 feet. Since one of these units was employed, it was necessary to place the control equipment in close proximity to the r-f equipment. However, because it was inconvenient to construct a shack close enough to the r-f equipment to permit use of such a short interconnection, the r-f equipment was separated from the antenna using transducers and coaxial cable, thus permitting an additional 30 feet of separation. See Figures 13 and 15.

The location for associated microwave control equipment also had to be satisfactory for locating the amplifying equipment for the eight wire channels. This then would permit one centralized location within the Amphitheatre where complete control could be obtained over the 19 video channels. The location selected was a large room immediately below the microwave roof location. Figure 16 shows a block diagram of the equipment layout in the room. The microwave equipment was divided so that the TE equipment was located on one side and the RCA on an adjacent side.

Each oscilloscope was associated with an RCA microwave receiver output, see Figure 17. Portable scopes not shown in the diagram were used for connection as required to the TE outputs. The *197* coils used for the wire circuits were mounted on a patch bay; see Figure 12. This patch bay was the central control point for all of our channels at the Amphitheatre. Electrically, it was located between the broadcasters' and our equipment and was designed primarily to permit a rapid switching of circuits or equipment in the event of trouble or changes in requirements. Twenty-five flexible coaxial leads were used to connect the various broadcasters' master control rooms on the second floor to this bay. The input and output of all equipment located in the room was connected to the patch bay with provision for monitoring at every point. This facilitated quick isolation of trouble.

At the Merchandise Mart, Figure 19, a special platform was constructed incorporating the same features for mounting the TE equipment as those used at the Amphitheatre. TE equipment was assigned inasmuch as the broadcaster used RCA equipment at this location which operated adjacent to our RCA band. It was felt that our use of RCA equipment might result in the introduction of mutual interference, especially during periods of tuning when high modulation levels were employed.

A power and ventilation room, Figure 20, immediately north of the platform was used to house the control equipment. From this location, coaxial cable was run to the *WNBQ* studios located on the floor below.

Three TE equipments were chosen for installation at the Civic Opera Building, Figure 21, for reasons outlined previously. Because of the substantial parapet height, a special raised platform was wedged between the 44th floor parapet wall and the building to provide equipment mounting space with sufficient clearance for the parabolas.

The control equipment was installed on the 45th floor just 150 feet north of this location. From this point connection was established to the studios of *WENR-TV* and the transmitter of *WNBQ* using short lengths of coaxial cable to the 44th floor and 42nd floor respectively.

The southernmost portion of the promenade roof on the 21st floor of the Hilton Hotel, Figure 22, was selected as a location for three RCA microwave equipments. A platform was erected to provide parapet clearance for the parabolas. A room just beneath the roof was selected to house the associated microwave control equipment. Since a substantial number of wire and radio channels originated and terminated here, the need for a control location was indicated. Therefore, the telephone equipment room on the 4th floor, Figure 3, was selected for installation of a patch bay. A balanced cable was employed between the equipment room and the roof location to provide a connection to the microwave equipment. Short lengths of coaxial cable were run from the patch bay to various broadcaster locations in the 4th floor and 5th floor.

PERSONNEL

The quantity of equipment planned for use during the political conventions required an increase in the number of personnel employed for television maintenance and installation. The normal force of 15 men was increased to 45 during this period with a proportionate increase in supervision. Because most of these new personnel had no previous television background it was necessary to establish a training course to include both theoretical and practical aspects of television installation and maintenance. This training period was begun some 2 months prior to the convention and was of approximately 4 weeks' duration. This permitted all personnel to have some practical training on the equipment before the convention started.

Design

The description given in the preceding paragraphs generally applied to preparatory work such as cable installation, equipment design, and so forth, based on estimates of circuit requirements received from the broadcasters. As actual orders for the channels were received, detailed circuit design was begun. Figures 1 and 23 depict the final layout for both microwave and wire channels. It should be emphasized that only channels ordered specifically for the convention are included in these layouts. Thus Figure 1 does not include, for purposes of simplicity, all of the channels available to each broadcaster which were in regular use during the convention period. This would include, for instance, in the case of *WNBQ*, six additional video channels from the Merchandise Mart used for regular network loops, remote studio pickups, and studio transmitter links, and in the case of other broadcasters a proportionate number of similar channels. Microwave facilities in normal use are also not included. As is evident, wire facilities extending to the Amphitheatre all employed equipment for transmission from this location. This was done to facilitate quick rearrangement of channels in the event of trouble and to simplify equalization and maintenance problems.

In addition, it was desired that each network feed from the Amphitheatre be provided with uniform facilities.

An extremely important consideration in establishing both wire and microwave channels was the provision of spare facilities. The quantity of channels which would be used during the convention was unprecedented in our previous experience and indicated that spare facilities should be available wherever possible. (These facilities were employed for our own use, however, and were not assigned to any particular broadcaster.) A study of Figures 1 and 23 will indicate that spare facilities were available in both directions between the Amphitheatre and Toll Number 2, between the Hilton Hotel and the Television Control Center in one direction, and between the Civic Opera Building (either ABC or NBC) and the Television Control Center in one direction. Additional facilities were available between the TCC and Toll Number 2 in both directions. Not included in the attached drawings are channels available from the TCC to each broadcaster for normal program use which could be used as additional paths during the convention period. Spare channels were equalized to each patch point so that many channel rearrangements were possible in the event of a failure of a particular section. To give an example, if one of the microwave circuits failed between the Hilton Hotel and the Amphitheatre, a spare wire channel between the Hilton Hotel and the TCC could be patched to a spare microwave channel between the TCC and the Amphitheatre to re-establish the circuit. This could be done in a matter of seconds in the event of a failure.

All channels, however, were not provided with spare facilities as is evident in Figures 1 and 23. The cost incurred in providing a complete spare system would be prohibitive and could not be justified on the basis of future usage.

MICROWAVE CHANNELS

Several problems were presented in the design of the microwave channels. In regard to the use of frequencies for operation of the RCA microwave units, our license permitted the use of only three frequencies for broadcast pickup purposes. Since five RCA units were planned for use on essentially parallel air paths to the Amphitheatre and others were in use on air paths likely to present interference problems, special temporary authority was obtained from the Federal Communications Commission to use frequencies licensed for closed channel operation (Theater TV). To reduce further the possibilities of mutual interference on adjacent channels operating over these parallel air paths, special twisted wave guide feeds were employed to rotate the polarization of the microwave energy by 90 degrees; see Figure 15, upper left. This virtually eliminated coupling with other units operating with normal polarization. These feeds were employed on equipment on the center frequency of the broadcast microwave band and on one of the low-band (Theater TV) frequencies. They were particularly useful during line-up periods when modulation levels were somewhat higher than normal.

Our TE equipment license permitted operation in the 3,700 to 4,200-megacycle band. This apparently large band available for selection of six channels of 20 megacycles width was limited by a number of factors. First of all, the transcontinental *TD2* microwave system extending east to New York and west to San Francisco employed frequencies in this range. The Loop buildings on which the TE microwave equipments were located were in proximity to the *TD2* antennas on the Toll Number 2 office. Since the orientation of the TE antennas was similar to that of the *TD* antennas, mutual cross-channel interference would result if common frequencies were employed. This limitation was further aggravated by the transmitter Klystrons used for the TE equipment. Only two of the six TE type equipments had transmitter Klystron tube mounts which permitted the use of a tube, recently made available, which would cover the entire 3,700 to 4,200-megacycle band. The old-type Klystron covered the range of 3,900 to 4,200 megacycles. Finally, only two of the microwave receivers contained wave guide filters which would suppress image frequencies. By the time we were able to squeeze a total of 120 useful megacycles out of this 500-megacycle band we felt very sympathetic to the Federal Communications Commission's problem in allocating frequencies.

It will be noticed in Figure 16 that a *1B* clamper amplifier was connected to the output of each microwave receiver at the Amphitheatre. Each unit performed two functions: first, to reinsert the d-c level before distributing the signal to the broadcaster; second, to permit an increase in level in the event of a decrease in the signal strength in some portion of the microwave circuit. The latter was necessary since even with optimum modulation the maximum output of the RCA or TE receiver into coaxial cable is approximately 1.5 volts peak to peak. The standard level used during the convention period in all broadcaster connections was 1.4 volts peak to peak resulting in what was thought to be a low gain margin.

WIRE CHANNELS

Little trouble was experienced in establishing wire line channels employing *A-2* amplifiers and other equipment which had been in use for some time. However, the new *A-2X* amplifier installations introduced many problems not encountered in preliminary development work. One involved the reduction of feedback between various points in the 85-decibel amplifier. This was solved by increased spacing between input and output leads, and by redesigning patch points and equipment bay wiring.

Another problem resulted from an attempt to equalize the six video wire channels extending from Wabash central office through Amplifier Station Number 1 to the Amphitheatre. To explain this problem additional *A-2* amplifier characteristics will be discussed. The *A-2* system originally was designed for a maximum equalization of 65 decibels due to the gain limitations of the associated amplifier equipment, that is, the equalizer inserts a frequency selective loss which makes the over-all loss of the cable plus equalizer uniform. This loss then must be made up by flat gain amplifying equipment to sustain the transmission level of a video signal. This 65-decibel limit postulates the use of a certain number of equalizers which are, within limits, not frequency selective below 10 kc. The cable attenuation loss, which increases with frequency when added to the reflection loss of the line which decreases with frequency, produces an approximately uniform response below 10 kc. Reflection loss arises from the fact that the source and terminating impedance of the *A-2* amplifier is 110 ohms, differing markedly from the characteristic impedance of the PSV cable below 10 kc. (As frequency increases the characteristic impedance of the cable starts from a large value with a highly reactive component and approaches a purely resistive value of 120 ohms.) The response, however, is not sufficiently uniform, requiring that each equalizer incorporate a slight correction. These corrections add to the proper value for 65 decibels or less but, with the large number of equalizers used in an 85-decibel line, an overcorrection results, producing a low frequency slope. All of the lines established through the repeater hut required the use of special equalizers to correct this low frequency slope loss.

Operation

Most troubles occurring during the period of both conventions were of an ordinary nature such as vacuum-tube failures, power failure, and the like. However, one particularly aggravating trouble involved equipment grounding at the Amphitheatre. This was evidenced as a 60-cycle signal superimposed upon the video signal and was objectionable only in so far as additional equipment was required. However, considerable apprehension was felt lest the interference become great enough to affect proper clamping. Without going into the many details of time-consuming efforts used to reduce this interference, it is sufficient to say that the trouble arose because of the differences in ground potentials between our location and the broadcasters' location. It does seem imperative in the future that installations of this kind should be planned in advance by the various organizations concerned and that methods of establishing a suitable ground system be agreed upon.

During the convention period a number of requests were received for channels to miscellaneous locations such as the Municipal Airport, hotels, depots, and so forth, with very little advance notice, requiring close co-ordination of commercial, engineering, and installation departments. In one instance, a request was received one evening for a circuit from the airport for the pickup of the arrival of President Truman's plane, the *Independence*. A circuit to this location had been established previously on several occasions using microwave equipment. However, recent construction work on a new hanger on the airport grounds had blocked the normal microwave path to the Loop area. Therefore, a new route had to be cleared quickly. The mobile television truck was maneuvered to a point on the field where a line of sight path was available and the microwave equipment was placed on the top of the truck. Transmission was good but was subject to interruptions by planes taxiing across the path. A conference with ground crew officials at the airport assured us that the path would be kept clear even though it meant some last-minute changes in their plans for "spotting" the *Independence*.

During the conventions, all responsible groups were organized to take instant action to minimize service outages which would naturally eventuate from the use of so much electronic equipment. However, months of preparation, planning, and training permitted us to reduce these outages to an almost negligible value. Of approximately 400 channel hours of program time which were provided to telecasters, only $9^3/_4$ minutes were lost, a percentage of 0.04. Program time is defined in terms of intervals during which the channel intelligence is actually being telecast. This does not include lost channel time, another index used to gauge our performance, which included failures occurring during the period the channels are ordered by the customer for line-up or stand-by regardless of telecast. This amounted to some 10 hours of a 4,000-hour total. Nevertheless, judging by public reaction to the national conventions there were some viewers who might have preferred to see a less impressive record.

No Discussion

Capacitors in Power Systems with Rectifier Loads

A. SCHMIDT, JR.
FELLOW AIEE

STATIC capacitors have come into widespread use as an efficient and economical source of reactive power for a-c systems. Harmonic voltages are sometimes a problem when applying capacitors to a system.

Years ago the principal sources of system harmonics were rotating machines and transformers. More recently, the widespread application of apparatus taking nonsinusoidal current, particularly mercury-arc rectifiers, has introduced other sources of system harmonics. These harmonics often require consideration when rectifiers and capacitors are connected to the same system.

Equivalent Circuit

If the voltage of a generator has a harmonic content of appreciable magnitude and if a capacitor and parallel inductance are connected to the generator, having equal impedances at one of the harmonic frequencies, harmonic currents of considerable magnitude will appear in the capacitor. Now consider the case of a harmonic-free generator connected to devices which take nonsinusoidal current from the generator. Such devices include rectifiers, transformers (whose magnetizing current is nonsinusoidal), certain types of electronic welding equipment, apparatus controlled by saturable reactors, and so forth.

Treatment of such circuits is facilitated by considering the devices as harmonic generators which deliver into the system the same harmonic currents which are normally taken from the system by the devices in question. This procedure will yield results in circuit problems which in many cases are only approximations, because the harmonic voltages resulting from the circuit action may react on the pseudoharmonic generator and modify its output.

Figure 1 shows how parallel resonance can occur when capacitors are connected to a system in parallel with a device taking nonsinusoidal currents from the system. Figure 1(A) shows a rectifier connected to a feeder fed by a step-down transformer from a large source. A power factor correcting capacitor is connected to the feeder near the rectifier. Figure 1(B) shows the equivalent diagram. The rectifier is replaced by a constant current harmonic generator having an internal resistance R_0, which is determined by the rating of the rectifier transformer. The feeder and step-down transformer are represented by an equivalent inductance in parallel with the capacitor. If the reactances of the capacitor and the

---

Paper 53-24, recommended by the AIEE Electronic Power Converters Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted October 21, 1952; made available for printing November 14, 1952.

A. Schmidt, Jr., is with General Electric Company, Schenectady, N. Y.

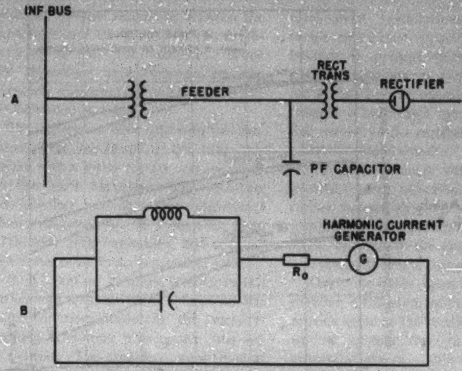


Figure 1(A). Typical circuit diagram of rectifier and capacitor connected to a-c feeder. (B). Equivalent circuit for harmonic analysis

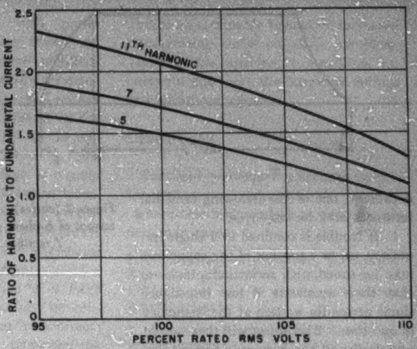


Figure 2. Harmonic current capabilities of capacitors with a single harmonic present

inductance are nearly equal at one of the frequencies of the harmonic generator, the currents in the capacitor and in the line will be substantially higher than the current delivered by the harmonic generator. The harmonic current may be high enough to damage the capacitor. There will be a high harmonic voltage on the feeder. This may cause local telephone interference. It may also increase heating of rotating machines on the system.

Rating of Capacitors

It is common practice to build power factor correcting capacitors with a kilovar capability which is 135 per cent of the name-plate rating with sinusoidal applied voltage. This permits a theoretical sustained overvoltage of 16 per cent without damage to the capacitor and constitutes a safety factor to provide for high line voltage and/or the presence of system harmonic voltages. Since harmonic voltages are present in any practical system, industry practice has been established which limits permissible maximum capacitor voltage to 110 per cent of rated voltage.

If the line voltage is maintained at rated value, the capacitor is able to handle harmonic currents of considerable magnitude, because the kilovolt-amperes due to harmonic current of a specified value are considerably less than the fundamental kilovars. For instance, a 7th harmonic current of 150 per cent of the rated fundamental current will result in a 7th harmonic voltage which is 21.4 per cent of the fundamental rating and 7th harmonic kilovolt-amperes of 32 per cent of the capacitor kilovolt-ampere rating. If there is no other harmonic present, the capacitor will be loaded to 132 per cent of its name-plate kilovolt-ampere rating even though the effective current is 180 per cent of the rated fundamental current.

Figure 2 shows the harmonic current capabilities of a capacitor which is loaded to 135 per cent of its kilovar rating, and shows how these capabilities are reduced when the fundamental voltage exceeds the rated value. If the total capacitor kilovolt-ampere due to the fundamental and harmonic current present in the capacitor exceed 135 per cent of the capacitor rating overheating and premature failure of the capacitor will result.

When capacitors are provided with protective devices, these devices normally respond to effective current. Thus, the capacitors may be removed from the circuit due to indication of overload, although actual overload may not be present.

A-C Harmonics in Rectifier Circuits

Figure 3 shows the form of the anode current in a 6-phase double-Y rectifier. The transformer primary current will consist of two such waves which have opposite polarity and are displaced 180 degrees. The latter current contains harmonics having orders of $6n \pm 1$, where n is any integer. If two or more 6-phase rectifiers are combined so that there are q phases, then the resulting harmonics in the a-c system have the orders of $qn \pm 1$.

Figure 4 shows the magnitude of the lower harmonics of a 6-phase rectifier as a function of the commutating angle u. The relation between commutating angle and commutating reactance is also shown. For a constant value of commutating angle, the magnitude of the harmonic is independent of the amount of phase control for values of over about 5 per cent.

The curves in Figure 4 are derived by rectifier circuit analysis, using standard conditions.[1] When capacitors which produce resonant conditions are introduced into the circuit, rigorous analysis is no longer practicable. Since it is convenient to use the concept of a harmonic generator in place of the rectifier, early investigators[2] working on interference problems employed a fictitious resistor in series with the pseudoharmonic current generator and then hypothesized a harmonic voltage generator having voltages which would deliver the currents in Figure 4 into the fictitious resistor. An empirically determined value for R_0 of $R_0 = \frac{1}{9} \frac{\text{transformer kilovolt-amperes}}{(\text{line amperes})^2}$ was found to give good agreement between observed data and calculations made according to this procedure when complex a-c networks were dealt with.

Remedial Measures

Because of the low amplitude of the higher order rectifier harmonics, capacitor overloading due to parallel resonance is unlikely if the kilovars of installed capacitors are less than about 1/2 of 1 per cent of the system short-circuit kilovolt-ampere at the point of installation. The possibility of telephone interference is present in this and smaller installations, and a survey is advisable when capacitors and rectifiers both having large kilovolt-ampere ratings are to be located in proximity to each other. If parallel resonance

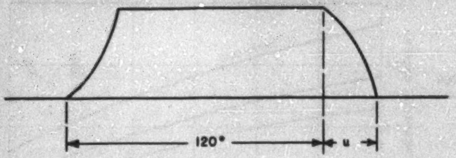


Figure 3 (above). Anode current of 6-phase double-Y rectifier

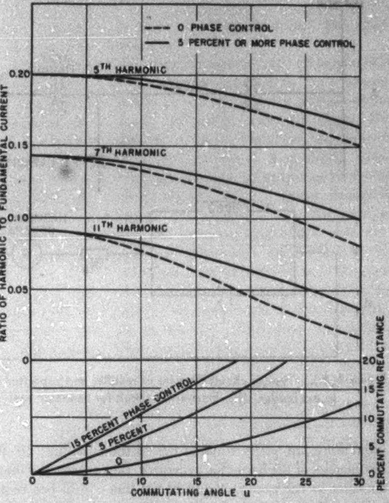


Figure 4 (right). Low order harmonic currents in 6-phase rectifier

is present, one of the following remedial measures may be employed:

1. If trouble is confined to a single frequency, all or a portion of the capacitors may be tuned with series inductance so that they constitute a low impedance shunt across the system at the frequency in question. This shunt will then accept practically all the harmonic current due to the rectifier at the resonant frequency, with a resulting system harmonic voltage determined by the impedance of the tuned shunt, which is low compared with the impedance of the capacitors. This is similar to the practice of using tuned shunts across a-c systems when rectifiers produce telephone interference due to the action between the harmonic currents generated by the rectifiers and the inductive impedance of the supply system. In such cases, shunts tuned to a number of the troublesome frequencies are connected across each phase of the supply system.

When a tuned shunt is employed, the harmonic current in the capacitor will be less than in the case of parallel resonance. Even so, it is possible to overload the capacitors in a tuned shunt, particularly when there are other sources of harmonic current present which add to the rectifier harmonic current. Thus care is needed in designing tuned shunts to assure sufficient capability on the part of the capacitors for all the available harmonic current at the resonant frequency.

If more than one harmonic frequency is present, a single tuned shunt may not eliminate harmonic troubles. The impedance of the tuned shunt will be inductive at frequencies higher than that for which the shunt is tuned. If all the capacitors are not used for this shunt, the remainder of the capacitors will be in resonance with a new value of inductive impedance at some higher frequency and difficulty may continue.

2. If series reactors are connected to all the capacitors to resonate at a frequency lower than any harmonic due to the rectifiers, then the series combination will be inductive at all frequencies due to the rectifiers and no resonant condition can exist between the capacitors and the system.

Any tuning arrangement will increase the fundamental voltage across the capacitors. For instance, in applying the foregoing scheme, if the lowest harmonic frequency in the system is the 5th, and the capacitors are tuned to the 4th harmonic, the resulting fundamental voltage across the capacitors will be increased from 100 per cent to 106.3 per cent of the line voltage. There will also be some 5th harmonic current since the tuned capacitors and the system will divide the 5th harmonic current in inverse proportion to their impedances at this frequency. The use of any tuning scheme therefore requires examination of the increase in capacitor loading.

In occasional applications, the need for series reactors with the capacitors may be anticipated. In such cases the capacitor designer can provide the higher voltage rating which is needed.

3. Sometimes resonance conditions occur under circumstances where no interference results and where the capacitors are not overloaded, but the capacitor fuses or other protective devices open the capacitor circuit because of the high effective current. In such cases, the level of current protection may be raised if it can be established that the total kilovolt-amperes permitted in the capacitors are within their rating.

In the majority of cases, capacitor protection is provided by fuses. These perform two functions: overload protection, and fault protection. In cases of a capacitor fault, the fuse should open before the capacitor case bursts. Co-ordination of fuse opening time and capacitor bursting time under fault conditions will establish the maximum fuse rating that may be used.

Case Histories

1. A 300-kw 6-phase rectifier was connected to a high-voltage system through a step-down transformer and a 4,000-volt overhead line approximately 4 miles in length. To correct the power factor of the load at the rectifier location, a 135-kilovar power factor correcting capacitor bank was connected to the 4,000-volt line near the rectifier.

During operation of the rectifier, there was frequent blowing of the fuses connected in series with the capacitors. Serious telephone interference was experienced. Analysis of the system indicated that the reactance of the step-down transformer and overhead line was approximately equal to the reactance of the capacitors at the 6th harmonic. The capacitors were sectionalized and inductances were connected in series with them to form 5th and 7th harmonic traps. Telephone interference and the blowing of fuses ceased.

2. Power factor correcting capacitors totaling 2,340 kilovars were located near an electrochemical installation having 4,300 kva of 6-phase rectifiers. Very serious telephone interference was experienced, centering about the 11th harmonic. Analysis of system impedance

indicated parallel resonance between the capacitors and the system in the neighborhood of the 11th harmonic. When the capacitors were disconnected, the harmonic voltages resulting from rectifier action became normal.

The capacitors were split into two sections. One was tuned to the 12th harmonic with a series reactor and the other to the 18th harmonic. The resulting combination provided low impedance shunts across the system to the 11th, 13th, 17th, and 19th harmonics, and reduced telephone noise to an acceptable value.

3. A 5,000-kw 12-phase rectifier installation was operated with a 1,125-kilovar capacitor installation in the vicinity. Serious telephone interference was experienced. The capacitors were fused at 1.8 times their rating and the fuses were often blown.

Investigation showed that the capacitors were in resonance with the system at a frequency slightly below the 11th harmonic. The 11th harmonic current in the capacitors was found to be substantially greater than the fundamental current and many times the 11th harmonic current in the rectifier supply, showing the effect of parallel resonance. Operation of the system with the rectifiers disconnected demonstrated that the rectifiers were the source of the 11th harmonic. Thus, tuning of the capacitors to the 11th harmonic could be depended on to reduce telephone interference to an acceptable level and to prevent blowing of capacitor fuses due to parallel resonance between the capacitors and the system.

4. Instances have arisen in large electrochemical installations with power factor correcting capacitors where unbalance is present between rectifiers. Such installations are generally arranged for 36 to 72-phase operation and no low frequency harmonics are present in the line current to the rectifiers if all the rectifiers share load equally. However, some 5th or 7th harmonic voltage is usually present because of action between transformer magnetizing current and system impedance. The presence of capacitors is likely to increase the system impedance and thus accentuate this action.

Thirty-six-phase operation of a large rectifier installation is obtained by connecting some of the rectifier transformers to the system through phase-shifting autotransformers. This causes the harmonic voltages to have a different phase relation to the fundamental in each rectifier. The output voltage of a rectifier is determined by the average voltage of an anode during its conducting period. If a harmonic voltage of the *n*th order is present, it will affect the average anode voltage unless the conducting period is an exact multiple of $360/n$ degrees, which is normally not the case. Furthermore, the phase angle of the harmonic voltage will influence the commutating angle, which in turn will affect the output voltage.

Thus, the presence of harmonic voltage will cause unbalanced load among the rectifiers. This in turn will increase the low frequency harmonic currents in the system so that a cumulative effect is produced with large resultant unbalance in load. This can be overcome by the use of phase retard on some of the rectifiers. This is objectionable because of increased commutating duty on these rectifiers. Tuning of the capacitors to a suitable frequency would limit the harmonic voltage and reduce unbalance.

Conclusions

Parallel-type resonance between capacitors and system reactance in the presence of harmonic sources such as power rectifiers may cause:

1. Overloading of capacitors, particularly if installed kilovars are much in excess of 1/2 of 1 per cent of system short-circuit kilovolt-amperes.
2. Operation of capacitor overcurrent protective devices.
3. Telephone interference.

These conditions can be remedied by the connection of suitable reactors in series with the capacitors when the voltage rating of the capacitors is not exceeded by so doing.

References

1. ANALYSIS OF RECTIFIER CIRCUITS, **E. F. Christensen, C. H. Willis, C. C. Herskind.** *AIEE Transactions*, volume 63, 1944, pages 1048–58.

2. EFFECTS OF RECTIFIERS ON SYSTEM WAVE SHAPE, **P. W. Blye, H. E. Kent.** *AIEE Transactions (Electrical Engineering)*, volume 53, January 1934, pages 54–63.

3. INDUCTIVE CO-ORDINATION ASPECTS OF RECTIFIER INSTALLATIONS, **AIEE Committee Report.** *AIEE Transactions (Electrical Engineering)*, volume 65, July 1946, pages 417–36.

4. INDUCTIVE CO-ORDINATION ASPECTS OF POWER FACTOR CORRECTION CAPACITOR INSTALLATIONS, **Joint Subcommittee on Development and Research.** *Engineering Report Number 50, EEI-51-16*, Edison Electric Institute and Bell Telephone System, New York, N. Y., 1951.

No Discussion

A Subscriber Toll Dialing Tape Reader

W. H. BLASHFIELD
MEMBER AIEE

Synopsis: Low cost machine scans punched tape produced by customer-dialed long-distance calls and prints a ticket or punches a card for each call. Process includes automatic rate selection and charge computation.

TO IMPROVE service and avoid the unprofitable manual handling of short-haul toll calls, systems have been developed which enable the telephone user to dial numbers in nearby towns. In one such system, intended for small independent offices and described in an earlier paper,[1] the call record is a punched tape as illustrated in Figure 1. Because the tape is not easy to read by eye, an automatic "reader" is needed to present the information in a form more suitable for the telephone company's billing department.

Several factors are involved in choosing the type of presentation. For the small company, simple and inexpensive equipment is the first consideration. For larger companies, a record permitting efficient handling of a large volume of business is important. In any company, the presence of operator-written toll tickets makes it desirable that the tape-reader record be capable of being handled in the same manner as the written tickets.

The availability of an inexpensive printer resulted in the development of a reader producing printed tickets. The tickets are the same size as standard toll

---

Paper 53-112, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted November 5, 1952; made available for printing December 5, 1952.

W. H. BLASHFIELD is with the North Electric Manufacturing Company, Galion, Ohio.

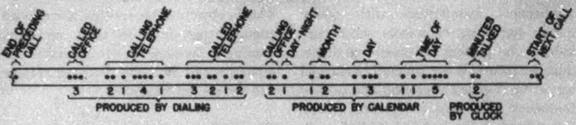


Figure 1. Punched tape. Called office and day-night entries are sometimes omitted

tickets, permitting them to be sorted and handled together with operator-written tickets. Furthermore, the electrical operation of the printer is similar to that of business-machine card punches, so that a standard card punch can be substituted for the printer if a company's volume of business is sufficient to warrant punched-card accounting.

The reader is shown in Figure 2, and a printed ticket in Figure 3. The punched black paper tape is received from the recording equipment on 8-millimeter movie reels, and is run through the reader backwards to avoid rereeling. A motor drives the tape past a phototube scanning unit which operates relay circuits that analyze the information and actuate the printer. Approximately 45 minutes are required to print tickets for one reel of tape carrying about 300 telephone calls.

The reader not only prints the information found on the tape but also computes and prints the charge for each call. If toll dialing is provided between two towns only, all calls take the same rate and computation is relatively simple. If a number of towns are involved in a dialing network, however, calls among several offices may appear on the same tape, requiring the reader to recognize the offices, choose the appropriate rate, and compute the charge based on that rate.

Because of the differences in scope of computation, two types of readers are made, one for the simple single-rate situation, and the other arranged for automatic rate selection. The simpler type contains about 10 tubes and 50 relays, the more elaborate version requires twice as many relays. The latter type will be described here.

Tape Information

To place an interexchange call, the subscriber dials an office code for the town he wishes to reach, then dials his own telephone number, and then dials the desired telephone in the distant office.

Office codes may be one or two digits, depending on the number of offices and arrangement of trunks in the network. With a 2-digit code, the first digit selects a tape recorder, and the second digit selects the desired office and also is punched on the tape to record the identity of that office. For example, if the subscriber dials office code *83*, the *3* is the first information on the tape, as illustrated in Figure 1.

The tape recorder consists of a punch magnet driven directly by dial impulses and a step magnet to step the tape each time it is punched. The step magnet can also be operated independently to provide separation between digits and numbers.

After the office code, the dialing of the subscriber's own number is recorded, followed by the dialing of the called number. (The customer is requested to dial his own number in order to avoid the complex and costly equipment that would be required to locate and record that number automatically. A simple circuit compares the number dialed with the calling line to insure that the calling number is dialed correctly.)

When the call is answered, the recorder engages a calendar, which punched a digit identifying the originating office, followed by a day-night digit indicating weekday or night-Sunday-holiday. Then comes the month, day, and time of day. (Time of day is in hour and tenth; thus *115* is 11:30 A.M.) Following the calendar information, a clock punches the tape once a minute to record the duration of conversation.

The day-night digit (*2* for weekday, *1* for night-Sunday-holiday) tells the reader whether day or night rates apply, night rates in telephone parlance being effective every night and also all day on Sundays and holidays. The calendar can supply day-night information to the tape rather easily, whereas it would be difficult for the reader to recognize Sundays and holidays from the date alone, although such readers have been built (Figure 2 is an example). As differentials between day and night rates usually exist only for distances greater than 40 or 50 miles, the day-night digit is omitted in many installations.

The called office digit also is absent in systems using single-digit office codes. This does not produce an ambiguous record, for in a single-digit system the recorders will be located in trunks directly connecting two offices, so that the calling office code punched by the calender is sufficient to identify both offices.

Ticket Information

The ticket of Figure 3 corresponds to the tape of Figure 1. Ticket information differs from the tape in that

1. the charge is computed and printed;
2. an *8* (or other appropriate digit) precedes the calling and called office digits to produce the complete office codes;

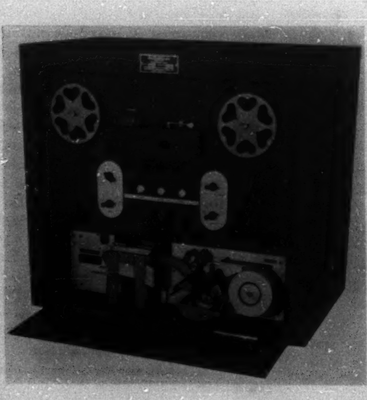


Figure 2 (left). Front view of tape reader with hinged bottom panel open to show printer

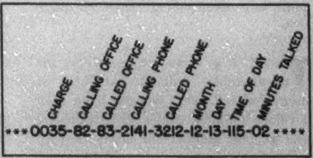


Figure 3 (right). Printed ticket

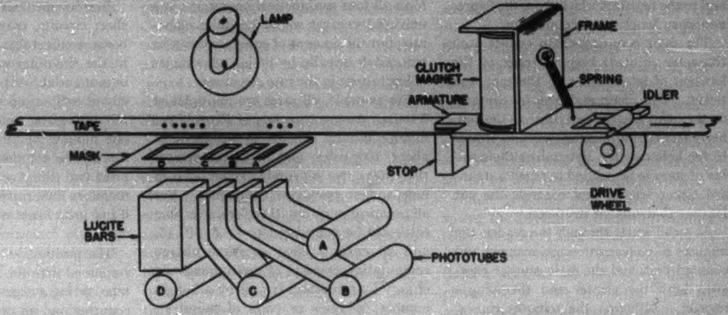


Figure 4. Scanner and tape drive principles

3. the calling office code is shifted to a position near the beginning of the ticket; and
4. the day-night digit is not printed.

The position of the calling office code is changed simply to make the ticket easier to read. The day-night digit is not printed because it only supplies information to the reader and would serve no useful purpose on the ticket. The charge appears at the beginning of the ticket because that is actually the end of the printed information with the tape and printer running backwards.

Ticket Printer

The printer is a *JP-2* Message Ticketer manufactured by the Teletype Corporation of Chicago, Ill.[2] A type wheel carrying the ten digits and dash and star is rotated continuously by a motor and friction clutch. On the same shaft is a pair of brushes rotating over a stationary commutator on which each segment corresponds to a character on the type wheel. If voltage is applied to a segment, a print magnet is energized when the brush reaches that segment, forcing the paper against the character on the type wheel. When the print magnet is released by de-energizing the commutator segment, a ratchet mechanism steps the paper ahead. The paper is on a continuous roll, and is cut into tickets by a magnetically operated knife.

Because the paper can be stepped only by operating the print magnet, dashes are printed to separate different numbers, and stars are printed to build out the ticket to standard toll-ticket length.

Printer operation consists of energizing a lead (commutator segment) corresponding to the digit to be printed; standard card punches are actuated in the same manner. For that reason, circuits designed to control the printer can also be used to produce punched cards.

Tape Scanner

Reader operation consists basically of counting each series of holes on the tape and then energizing the printer lead corresponding to the digit counted. Series of holes are recognized by the spaces between them.

Three widths of spaces are used: (1) narrow between digits, (2) medium between numbers, (3) wide between calls. Any space signals the reader to print the digit counted: a medium space indicates that a dash is to be printed in addition to the digit, a wide space indicates the end of the call information.

The scanner and tape drive mechanism are illustrated in Figure 4. The tape rides over a mask containing four slots, lettered *A*, *B*, *C*, and *D*. Lucite bars guide the light from each slot to a corresponding phototube. Narrow slot *A* is used for hole counting, impulses from phototube *A* driving a relay counting chain. Slots *B*, *C*, and *D* are for space recognition, slot *B* being covered by a narrow space, *B* and *C* by a medium space, and slots *B*, *C*, and *D* by a wide space.

A clutch controls the tape drive. When the clutch magnet is energized, the left-hand end of the armature is lifted away from the stop against which it has been holding the tape, and the idler at the opposite end is forced against the continuously rotating drive wheel to drive the tape. After each series of holes scanned, the space covering slot *B* releases the clutch and closes a circuit to print the digit that has been counted. After printing, the counting circuit is reset and the clutch re-energized to start the tape for the next digit.

As mentioned earlier, the tape is scanned backwards to avoid the need for rereeling. Backward operation has the further advantage that incomplete calls can be recognized and ignored.

From the tape of Figure 1, it will be seen that the first information scanned is the minutes of conversation, followed by a medium space. If, however, the called number had been busy or had not answered, no calendar or clock information would have been punched in the tape; then the first thing scanned would have been the last digit of the called number, followed by a narrow space. If the reader sees a narrow space after the first digit scanned, it prints nothing, but simply lets the tape run until the next call is reached.

Block Diagram

Referring to Figure 5, the tape scanner drives the hole counting chain which furnishes information both to the printer

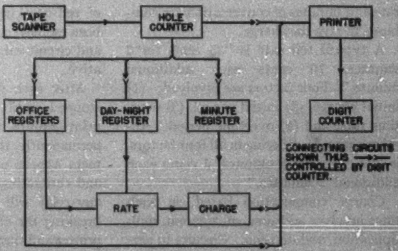


Figure 5. Block diagram of reader

and to the registers which store the figures necessary for charge computation.

The digit counting chain counts each character printed, keeping track of the location of information on the tape and ticket. This chain is driven by contacts on the print magnet and controls the connection of the various registers and printer to the hole counter, determines the number of stars to be printed to reach a standard ticket length, and controls the cut magnet that cuts off the ticket.

As a call starts through the reader, the minutes of conversation are scanned and counted first, and the digit counter energizes both the printer and the minute register. Although the minutes consist of a single series of holes, the counting circuit breaks the information into tens and units digits.

Next the time of day, day, and month are scanned and printed, but not registered. Then the day-night digit and the calling office digit are scanned and registered, but not printed. The digit counter controls the connection of the registers and printer at the proper points. When digits are registered without printing, the registers step the digit counter.

The called and calling telephone numbers are printed, and last comes the called office digit, followed by the wide space marking the end of the call. When that digit has been printed, the reader automatically prefixes an *8* to complete the office code. Then the printer is connected to the office register to print the calling exchange digit which had been stored earlier without printing. Here, again, an *8* or other appropriate digit is automatically prefixed.

At this point the printer is switched to the computer circuit, which delivers the charge for the call. Finally stars are printed until 40 characters are reached to give a standard length ticket.

Computing

As indicated by the block diagram, the combination of offices and the day-night indication determine the rate; then the rate and minutes of conversation are combined to give the charge.

A typical toll rate is "35 cents for 3 minutes, 10 cents each additional minute." Four factors are involved: (1) initial charge, (2) initial period, (3) overtime rate, and (4) overtime period. Another example, differing in all four factors, is "20 cents for 5 minutes, 5 cents each additional 2 minutes."

Charge computation based on such rates involves addition, subtraction, multiplication, and division. Circuits to perform all four arithmetical operations in a universal manner would be quite elaborate, but the amount of equipment can be reduced drastically by taking advantages of limitations in the rate structure.

For example, all rates are multiples of 5 cents. An overtime rate of 20 cents per minute usually applies to distances of about 100 miles, and distances greater than 100 miles are rarely encountered in independent telephone companies. For all practical purposes, therefore, overtime rates can be considered to be 5, 10, 15, and 20 cents. Then overtime charge computation consists of multiplying one of only four numbers (rates) by a 2-digit number (minutes or pairs of minutes). When limited to this extent, multiplication requires relatively little equipment.

Rates are set in the reader by strapping between terminals. Each office combination is represented by a set of four terminals. Another group of terminals represents rates and periods. Each of the four office-combination terminals is strapped to an appropriate rate or period terminal to give the initial period, initial charge, overtime rate, and period for that office combination.

If there is a differential between day and night rates, sets of day-night terminals are interposed in the strapping between the office-combination terminals and the rate-and-period terminals to give different rates for day and night.

The strapped terminals are on a multiconductor plug which fits a permanently wired jack in the back of the reader. Different plugs can be strapped for different rates and office codes; then if a large company has several separate groups of offices, the corresponding rate plug can be inserted in the jack when running tapes from a given group of offices.

Circuit Design

Reliability, simplicity, economy, flexibility, and speed were factors governing the reader circuit design. The importance of reliability and accuracy in a machine producing information for thousands of telephone bills is obvious. Various types of self-checking circuit principles have been employed, and choice of components and circuit values has been very conservative.

After each digit scanned, the tape is stopped until the printer prints; any print failure will leave the machine permanently stopped. Only make contacts appear on register, rate selection, and computing relays, so that failure of any relay will result in an open circuit, breaking the chain of information to the printer and thereby stopping the machine.

Besides open circuits, it is possible that short circuits, crosses, or grounds might occur accidentally. These usually result in the simultaneous energization of two or more print leads. For example, a short circuit or spurious relay operation in the minute register would look like more than one number of minutes, resulting in an attempt to compute two charges and to print two things at once. A trap circuit monitors the printer and stops the reader if two print leads are energized simultaneously.

The positions of the digit counter are compared with the medium spaces on the tape, giving a check both on correct digit counting and on proper number of digits per number on the tape.

Because the spaces on the tape are measured by physical distance (slots *B*, *C* and *D* in Figure 4) rather than by any method involving time, tape speed is unimportant. There is no lower limit on speed, for all tube circuits are direct-coupled and no slow relays are used. At the other extreme, readers are tested for satisfactory operation at 100 holes per second, or twice the normal tape speed.

Electron tubes are long life types operated conservatively, and the number of tubes and tube types has been held to a minimum. The light source is provided with efficient reflector and lens to produce intense illumination at the scanning mask, resulting in more than adequate output from the phototubes without the use of undesirably high load resistors. The black paper cuts off the light completely when spaces in the tape cover the slots.

Relays are standard telephone types of proved reliability in circuits which avoid marginal operations or close timing. Forced-sequence principles are used extensively, each step in the circuit operation depending on, and being initiated by, the completion of a preceding step. This provides maximum reliability, because a failure at any point stops the entire operation; it also gives maximum speed, because operations follow each other at their own inherent speeds rather than being assigned definite times in which to occur.

Simplicity is of great importance in equipment intended for small independent telephone companies, for such companies rarely have skilled men capable of understanding and maintaining complex equipment. Unfortunately, a machine which performs all the functions of the tape reader cannot avoid a certain amount of complexity, but circuits have been kept as straightforward and understandable as possible, and the design has adhered to types of circuits and apparatus which are most apt to be familiar to the average telephone man.

The printer is rugged and quite simple mechanically, permitting adjustment or repair by a person with no unusual mechanical skill or training. The only other mechanical parts in the reader are the tape drive motor and clutch.

Relays have been used in preference to electron tubes wherever possible, because the small-company telephone man usually knows nothing about electronics but has at least a speaking acquaintance with relay circuits as the result of the extensive use of relays in dial switchboards.

Observation and testing are facilitated by test keys. The reader can be made to scan one digit and stop, print that digit and stop again, and so on, each successive operation being initiated by the manual manipulation of a key. This slow-motion principle is augmented by a switch to turn off the tape drive motor, permitting the tape to be fed very slowly by hand so that tube voltages and relay operations can be observed for each hole in the tape. With these facilities, a case of trouble can be isolated rather quickly and easily.

Low cost is facilitated in both models of readers by the use of a simple and inexpensive printer. Furthermore, the single-rate model avoids much of the relay equipment found in the automatic-rate-selection model, bringing the cost in range of the smallest telephone companies.

In the single-rate reader, the registers and rate and charge circuits of Figure 5 are eliminated, leaving only the tape scanner, printer, and hole and digit counters. With a single rate, the charge is entirely a function of the number of minutes talked; consequently the hole counting relays have contacts wired in such a manner that when minutes are counted, the corresponding charge is printed instead of the number of minutes. On the remainder of the ticket, the information is printed in exactly the same form as it is found on the tape. This elimination of all storage and separate computing circuits results in a reader with simple circuits and a minimum of equipment.

The single-rate model is not limited to systems containing only two offices, but rather to systems in which all calls on a given tape take the same rate. For example, a network might contain several towns but have the trunks and recorders arranged in such a manner that a given tape recorder handles calls between only two towns. When running tapes from that recorder, the hole counter in the reader will be wired for the corresponding rate; when running tapes from a different recorder, the hole counter wiring can be changed to a different rate by a multicontact switch or by removing and inserting different multicontact strapped plugs.

High speed of operation is desirable in readers used by larger telephone companies. This is primarily an economic factor, of course; if doubling the speed of the reader more than doubles the cost, it is cheaper to use two slow readers than one fast reader.

Rate of ticket production is determined primarily by the speed of the printer and the speed of counting holes in the tape. The optimum counting rate is about 50 to 75 holes per second; at any lower rate the printer has to wait for the counting, whereas higher counting speed does not appreciably increase ticket production because the printer is unable to go any faster.

For reliable counting at, say, 50 holes per second, the counting circuit should be capable of following 100 holes per second. This is rather fast for conventional relay circuits, yet it was desired to use relays instead of electronic countres for the sake of circuit simplicity and flexibility. If we follow the obvious procedure of operating a relay from phototube *A*, that relay operating and releasing from every hole in the tape and driving a relay counting chain, we are confronted not only by a problem of speed but by the problem of mechanical wear and contact erosion on the rapidly operating driving relay.

These problems were solved by eliminating the driving relay and developing a relay counting chain driven directly by an electronic circuit. This circuit is essentially a gas tube "flip-flop" controlled by phototube *A*, with the relay chain advanced one step by each reversal of the flip-flop. This scheme, besides eliminating the speed and wear problems of a driving relay, inherently doubles the speed at which the chain can count and actually increases the speed still further by eliminating problems of pulse ratio, thereby permitting counting at a speed of 100 per second with conventional telephone relays.

The reader normally drives the tape at 50 holes per second, at which speed 350 to 400 tickets per hour are printed, or about 3,000 tickets per 8-hour day. This is adequate for the toll traffic of small- and medium-sized independent telephone companies, and the low cost of the machine permits using two or more readers in companies with larger volumes of business.

References

1. **Simplified Subscriber Toll Dialing for Low-Rate Low-Density Toll Circuits Between Community Dial Offices, Eric Brooke, J. M. Blackhall, W. H. Blashfield.** *AIEE Transactions,* volume 69, part II, 1950, pages 1417–20.

2. **Automatic Ticketing, O. A. Friend.** *Bell Laboratories Record* (New York, N. Y.), volume 22, July 1944, pages 445–50.

No Discussion

Nuclear Power Plant Control Considerations

M. A. SCHULTZ
MEMBER AIEE

THE development procedure for designing the control system for a nuclear power plant is basically the same procedure as for any other type of power plant. The presence of a nuclear reactor as a component in the system imposes many unusual design problems. The purpose of this article is to review general plant considerations affecting control system design, giving special emphasis to the features of reactor control.

The basic method of approach to control system design consists first of studying the plant operational requirements and the plant control requirements. Following this review a preliminary block diagram of a control system to fulfill these requirements may be attempted. However, after a study of this diagram, it often becomes apparent that the inherent dynamics of the plant and the characteristics and limitations of some of its components are of utmost importance. Modifications are then made to the proposed block diagram. Once the desired operation of the control system is fully understood, the problem of synthesizing the control system can begin.

At the present state of the art the problems introduced by the reactor necessitate more study than do the conventional ones. Some literature has been written about reactor control,[1-4] but very little is available on the control of power plants employing a reactor. An outline of the considerations required before a plant control design can be completed is given here.

Plant Operational Requirements

Certain information must be known concerning the operation and output of the plant. These operational requirements are usually given in the basic plant specification.

For the sake of illustration an elementary block diagram of a possible nuclear power plant is shown in Figure 1. Here the reactor is considered merely as a source of heat energy. When this heat is extracted by passing a coolant through the reactor, the heat energy is then transferred via a heat exchanger to a turbine system, and the output shaft of the turbine can be made to drive many types of working apparatus. The primary coolant considerations are basic for control design, and it may be mentioned in passing that gas, water, and liquid metals have been proposed as reactor coolants.

The ultimate usage of the system of course determines the specified output requirements. A useful principle of control design is to attempt to place the primary operating controller near the output of the system. Consequently, the characteristics of the output loading as a function of turbine horsepower output is of first interest. In the case of an aircraft power plant this information might be specified in terms of aircraft speed as a function of turbine output. For a shipboard plant the information might be supplied in the form of propeller rpm versus turbine output. In any case, an understanding of the relationship between the turbine and the load output is necessary to plant control.

Another factor affecting the type of control system is the operational transitions required. These transitions determine the range of control and the output level variations required. Minor consideration is given to whether the final output is to be continuously variable, variable in steps, or some combination of the two.

Start-up and shut-down requirements affect the control system. Here the nuclear reactor plays a prominent part in determining plant performance, for nuclear reactors cannot be started up too fast and never can be really shut down. In considering the output of the nuclear reactor there is no such thing as zero power involved. Even in a brand-new, cold, clean reactor which has been shut down as much as possible an inherent source of neutrons exists which causes a certain amount of nuclear fission to occur. This fission produces minute amounts of power and in the process more neutrons are released. This effect requires that the reactor neutron output be monitored continually to see that the neutron level is kept under control. In practical cases this monitoring means that instrumentation for measuring neutrons must be capable of operating over ranges as wide as 6 to 10 decades and control provisions must be provided over a similar range. Starting up a reactor from a cold condition means that the reactor can operate over a neutron range of many decades before any appreciable useful power begins to occur at the output. Some arbitrary number of neutron fissions has to be taken as zero power. For example, 1 per cent of full-load out-put might be considered to be the equivalent of zero power. Transitions between 1 per cent full power and 100 per cent full power can be made relatively rapidly, but because of the safety aspects involved transitions from a cold start-up must be made quite slowly.

A reactor which has been operating at a high power level is turned off neutron-wise by manipulating the reactor control rods to reduce the reactivity of the reactor. However, many secondary radiations, particularly gamma rays, exist from the formed fission products. The amount of power from these secondary radiations is in some instances very substantial. Consequently, although the reactor has been turned off to the best of the ability of the operator, serious amounts of power still may be given off for a considerable length of time. Control-wise, this condition means auxiliary devices may be necessary to dump this spare power into a useless load. It can be seen, therefore, that the start-up and shut-down of a nuclear power plant is dependent upon the reactor characteristics and differs from conventional plant techniques.

The transient output requirements of the plant are also of some importance. From an operational point of view certain peak demands will be made on the system. These demands are similar to those made on a conventional plant whereby above rating performance can be obtained for short intervals of time. The operational probability of such demands and the operational probability of any preferential power level of operation must be analyzed. For example, in a central station plant where a relatively fixed load can be made to exist it might be desirable to operate the nuclear powered generator at 100-per-cent capacity during most of its lifetime.

Paper 53-31, recommended bythe AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted October 16, 1952; made available for printing November 14, 1952.

M. A. SCHULTZ is with the Westinghouse Electric Corporation, Atomic Power Division, Pittsburgh, Pa.

The work concerned in this report was carried out under Atomic Energy Commission Contract *AT-11-1-GEN-14*.

Acknowledgment is made of the assistance rendered by J. C. Simons, J. N. Grace, and G. H. Conley of the Westinghouse Atomic Power Division.

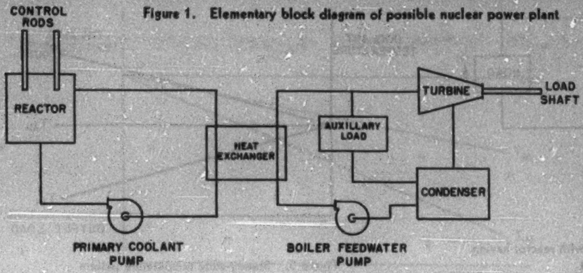


Figure 1. Elementary block diagram of possible nuclear power plant

If this were the case the plant control system might be optimized at this load value.

The design of the reactor with respect to poison overide is another important operational control consideration. When some types of reactors are shut down, nuclear poisons are known to build up and reduce the over-all reactivity of the reactor. In time these poisons die away. A certain amount of additional reactivity can be designed into a reactor to overide this poisoning effect. The amount of the poisoning overide which is built into the reactor becomes an operational factor because, with only a small amount of overide reactivity in the system, the poisons can build up in a given amount of time to an amount capable of making the reactor inoperative. Before this time is reached the reactor might be turned back on and the poisons effectively burned out. If, however, the reactor poisons have built up to such a level that the plant cannot be turned on, it then becomes necessary to wait until the poisons die down to where the reactor can again be made critical. To build enough reactivity into a reactor to always overide the poisons may require a large amount of fissionable material. If the service to which the power plant is put is such that the plant must be shut down frequently for critical periods then some other means must be found for supplying auxiliary power during the dead off time of the reactor.

Another of the factors which is involved in the design of control components is the reactor and plant lifetime. Contrary to popular opinion, reactors do not have an indefinite life. The uranium atoms which are involved in the fission process become used up and although a tremendous amount of energy is available each time a uranium atom is fissioned, ultimately enough fissions occur so that the reactor becomes depleted in uranium and will no longer maintain a chain reaction. Other factors, such as radiation damage to the structural material of the reactor or the corrosion of any portion of the mechanisms, may cause the end of life of the reactor. If the design of the plant can be made such that all components have an effective indefinite lifetime compared with the lifetime of the reactor, it may then be feasible to employ some means of changing the fuel or of possibly even changing reactors.

Plant Control Requirements

Once the over-all characteristics required of the plant have been specified, certain of the details of these characteristics pertaining directly to the control system can be examined. Detailed control information that is needed is the steady-state output error that can be tolerated. That is, how important is it that the plant give out a fixed number of kilowatts or a given propeller rpm and hold this output regardless of operational transients. The steady-state output error directly specifies the tolerable steady-state plant instrument errors. Frequently, control systems are devised whereby the errors of many instruments in many loops add up to produce a higher probable error than the over-all specifications permit. On the other hand, it is possible in the design of certain loops to integrate out instrument errors in such a manner that they do not show up in the final output. From the over-all error permitted the designer can work backwards through each loop, figuring the steady-state error that conventional control components will impose upon the system.

Similarly, certain plant and reactor transient errors may be permitted for some of the instruments. For example, if two thermometers are used in the system and one thermometer has a much larger time constant than the other, it is obvious that in a transient change both of these thermometers will not read identically. If these instruments are used in control loops the output of a particular loop may temporarily permit an error to exist in the plant output.

One control requirement which is peculiar to the nuclear power plant is the instrumentation setup for alarm, cutback, and scramming circuits. Alarm circuits are common in the power industry, but the peculiar notion of "scram" is one limited to the reactor field. A scram in reactor parlance consists of an attempt to insert as much negative reactivity as possible into the reactor in as short a time as possible. Practically, a scram usually means injecting safety rods into the reactor with some violence. The safety requirements of the reactor are such that it is usually necessary to employ scramming devices. However, from a practical power point of view it is obvious that the plant must be kept in an operable state as continuously as possible. The plant must not be shut off every time some minor control in a secondary loop malfunctions. Therefore the control designer's problem is to limit the number of controls which can cause a scram to an absolute minimum.

For all accidents other than the type requiring a scram, it is possible to cut back the reactor power level or to merely ring an alarm and have an operator manually make the necessary adjustment or shutdown.

Plant Dynamics

The heart of the control system problem is the primary coolant system operating cycle. There are many patterns into which the temperature and the flow of the coolant coming in and out of the reactor may be manipulated to produce the desired output power. The system operating cycle may be determined by many items such as auxiliary requirements, corrosion, or peak temperatures permitted. However, in a control system designed about a reactor one should attempt to follow this basic premise: If the reactor has a negative temperature coefficient the reactor system would prefer having the average coolant temperature constant. A word of definition is necessary concerning the reactor temperature coefficient. To say that a reactor possesses a negative temperature coefficient means that as the reactor average temperature increases its over-all reactivity decreases. That is, the chain reaction will slow down with increasing temperature. This condition is obviously desirable.

Let us examine the simple plant shown in Figure 2 in the light of this premise. This figure shows a primary coolant being circulated at constant flow through a re-

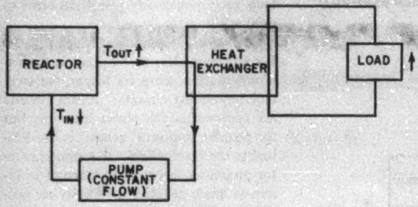


Figure 2. Simplified power plant block diagram with reactor having negative temperature coefficient

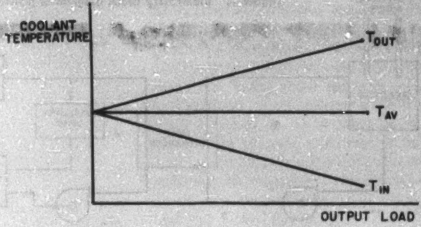


Figure 3. Steady-state temperature pattern

actor having a negative temperature coefficient. A steady output is being produced; now suppose that more output is required by the load. This greater loading causes more heat to be extracted from the heat exchanger and for a short period of time the heat capacity of the heat exchanger can usually supply this load. However, the additional energy extracted from the system requires that the temperature of the coolant into the reactor must drop. If the reactor has a negative temperature coefficient as defined here, the dropping of the coolant temperature inserts more reactivity into the reactor. If the reactor was initially in a critical state it now temporarily becomes supercritical. More energy is then available from the reactor, and consequently the output temperature of the coolant rises, and in the steady state the reactor returns to its critical condition with the average coolant temperature the same as it was initially. These steady-state temperature conditions are shown in Figure 3. It will be noted that without any control mechanism whatever the reactor system has stabilized itself about a given average temperature and automatically supplied any reasonable demand put upon it. This is the case only where the program of primary coolant operation is such that the average coolant temperature remains constant.

Let us assume now that for some reason other than the reactor control system it is desirable to establish a different pattern of coolant temperatures. A reasonable pattern might be that shown in Figure 5. This figure shows a simple system which might be called for by some structural condition desiring to keep the outlet temperature fixed at a given maximum level. Figure 4 shows a block diagram of how this condition might be achieved. Again if more load were required the inlet temperature to the reactor would drop. More reactivity would be inserted in the reactor because of the negative temperature coefficient. The outlet coolant temperature would tend to rise, but now a control has been inserted on the outlet temperature which measures the temperature and then varies the coolant pump speed in such a manner as to reduce this outlet temperature. In this manner it is very feasible to hold the outlet temperature constant as a function of load, as seen in Figure 5. However, it will be noted that the average coolant temperature drops with load, meaning that the reactor, because of its negative temperature coefficient, would tend to operate in a supercritical condition. Consequently another control loop, usually rod control, must be added to the system to extract the reactivity put in by the change in coolant temperature.

The system of Figure 4, from a control point of view, compares unfavorably with the previous system. Here the external control systems which have been added cause the reactor system to operate in a manner contrary to its own inherent stability. Consequently, the system is not quite as safe and requires considerably more control equipment. Neither of the simple systems described is completely practical for an actual plant, but many derivatives and variations can be derived which will fulfill the required specifications in a compromise manner.

Once the primary system operating program has been determined it would be very desirable to have the transfer functions of all of the major system components available. Unfortunately, in the present state of the art the control designer must be satisfied with far less. Usually a simplified approximation to the transfer function of a given component is made analytically, and the simplification is usually so great that the resulting equations can be handled easily in multiloop analogue networks.

Still more information is needed, however, before the designer can attempt to synthesize a complete control system. The transport times of the various coolants and the various loops must be known. Plants can be laid out with many feet of pipe between major components, and the time delay in these pipes is an important part of the control system. Of particular need is the total coolant circulating time, the time the primary coolant is in the reactor, the time the primary coolant is in the heat exchanger, and the time the primary coolant is in the pipes.

Miscellaneous Plant Information

There are many miscellaneous bits of information which also must be considered before the control system can be synthesized. Some of these factors depend upon the type of main coolant system used. For example, in systems employing water or gas as the coolant, it will usually be found necessary to pressurize the main coolant system in order to obtain reasonable plant efficiency. Operating a

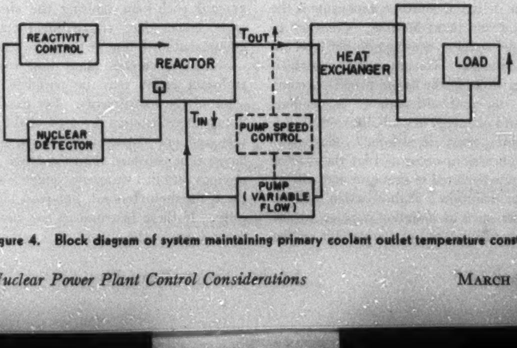


Figure 4. Block diagram of system maintaining primary coolant outlet temperature constant

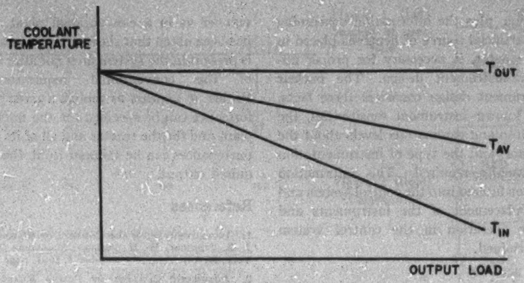


Figure 5. Steady-state temperature pattern

loop under pressure calls for a complete pressurizing system including heaters and controls. These controls of necessity tie in with the controls of the main coolant system.

Another auxiliary system which directly affects the main system is provision of devices to take care of emergency electric power failures. It may be desirable in some plants to provide auxiliary means of dumping power or employing elaborate fail safe mechanisms.

A third item which affects the design is the type of coolant control permitted. If pumps or blowers are used to circulate the primary coolant, multiple-unit operation may be considered. Multispeed units, variable speed units, or the switching of units on or off the line are methods whereby coolant control may be achieved. One scheme may have a considerable advantage over another in a particular plant design.

Another factor is the auxiliary load which the plant must supply in addition to its primary output. In a nuclear power plant for a vessel, for example, a considerable hotel load exists separate from the prime task of driving propellers. It may also be desirable that most of the direct auxiliaries, such as pumps, be tied on to the nuclear powered system. This statement implies that a certain percentage of the auxiliary load must be furnished from a source other than the nuclear power plant. Just as the battery on an automobile is required to run certain of the automobile's auxiliaries, a given amount of non-nuclear power is usually necessary to start the plant and keep it in operation. To start the plant and its auxiliaries and maintain them under given operating conditions usually will require a large amount of detailed electric switchgear. This apparatus, though complicated in its own right, is conventional in nature and is usually capable of supplying the control system designer exactly what he wants.

When all of these factors have been considered, the control system designer can then synthesize the entire system. The problems of the designer now involve problems of components, particularly the limitations of the various components. These limitations usually lead to the separate protection circuits for each of the major components.

Component Characteristics and Limitations

Components such as the turbine, the heat exchanger, the reactor vessel, the pumps, and the reactor itself require the most study. As each plant will contain a different group of components with different limitations, only a few of the more obvious conventional characteristics and limitations of the plant will be mentioned in order that more emphasis may be placed on the reactor characteristics.

Of interest to the control designer are the heat exchanger temperature limitations. The structural limitations of the heat exchanger will probably permit only certain rates of change of temperature during both start-up and shut-down operations. In other words, a heat exchanger may have a limitation stress-wise such that the temperature of the coolant flowing through it may be permitted to change only by so many degrees per minute. A completely different rating, however, may be given the boiler for very short time transients. The same condition exists with respect to the reactor container. This container will have stress limitations also which will prevent any attempt to change coolant temperatures faster than a given rate.

Minor component limitations exist in the piping and the specific instrumentation of each component. Consideration of these limitations usually lead to the settings on the relief valves.

Reactor Nuclear Characteristics

It is now necessary to examine roughly the various factors contained within the reactor itself which affect the over-all plant control. It has been shown[2] for control purposes that the reactor can be considered as a nonlinear device having an approximate transfer function of the form

$$K_rG_r(s)=\frac{K_rN_0}{s+a} \qquad (1)$$

where

K_r and a are constants
s is the Laplace transform operator
N_0 is the neutron level at which the reactor is operating

The problem of controlling a loop containing this component whose gain depends upon the level at which it operates has been given extensive study.[2,3] The ultimate objective of this study was to be able to specify components of a control system which would manipulate the nuclear control rods. In practice, the nuclear characteristics of the reactor are also involved in the detailed rod moving mechanism design.

To the component designer the problem of rod mechanism design is a simple one of merely asking how far a given control rod must move and how fast. These questions can be answered by the reactor designer who must determine first the total number of control rods required and where they are placed in the reactor. These answers are derived from heat transfer considerations. He must then consider the problem of the effectiveness of these rods in changing the over-all reactivity of the reactor. He must also consider how this effectiveness changes during the lifetime of the reactor. The variation of these constants with time brings up for consideration again the kinetic relationships of the various poison build-ups and burn-outs in the reactor. Reactor fuel depletion must be considered again and the total amount of reactivity to be built into the reactor determined.

The next reactor consideration is that of the magnitudes of the various coefficients of reactivity. The temperature coefficient of reactivity has been mentioned previously and its absolute magnitude is necessary to enable the determination of the degree of stability of the primary loop. Another coefficient of reactivity may be determined for some reactors and this is a pressure coefficient. This term is defined in a manner similar to the temperature coefficient and consists of the amount the reactivity of the reactor changes, and in what direction, as the pressure of the coolant in the reactor changes.

Another nuclear constant of importance to the control designer is the so-called "mean neutron lifetime." This is the average time taken by a neutron in starting out at birth in fission and dying by creating another fission or being absorbed or lost in some other portion of the reactor. The importance of the mean neutron lifetime control-wise comes about in that it affects the constant *a* of equation 1.

Start-up and shut-down reactor considerations call for knowing the amount of non-neutron power produced immediately after shutdown and the neutron level to which the reactor will ultimately drop to after a prolonged shutdown. This information, plus the information concerning the artificial source of neutrons placed in the reactor, is necessary for proper nuclear instrument design. The nuclear instrument design considers these facts, the known instrument sensitivities, the neutron and gamma flux levels about the reactor, and the type of instrumentation philosophy required. This information is then filtered into the control system and the placement of the instruments and their operation in the control system determined.

Conclusion

Many of the factors involved in the control of a nuclear power plant have been mentioned and it is apparent that the operational requirements of the plant determine the control system in the same manner as in a conventional plant. It has been noted that the fact that a reactor is present in the system does not alter any of the conventional requirements. Rather it appears as though a given performance can be specified for the nuclear plant and the the reactor and all of its appurtenances can be tailored to fit the required output.

References

1. INSTRUMENTATION AND CONTROL OF REACTORS, **J. D. Trimmer, W. H. Jordan.** *Nucleonics* (New York, N. Y.), volume 9, number 4, 1951, page 60.

2. AUTOMATIC CONTROL OF POWER REACTORS, **M. A. Schultz.** *AEC Document 3163*, Atomic Energy Commission, Washington, D. C., 1950.

3. CONTROLLING A POWER-PRODUCING NUCLEAR REACTOR, **J. M. Harrer.** *Nucleonics* (New York, N. Y.), volume 6, number 3, 1950, page 58.

4. DERIVATION AND INTEGRATION OF THE PILE-KINETIC EQUATIONS, **H. Hurwitz, Jr.** *Nucleonics* (New York, N. Y.), volume 5, number 1, 1949, page 61.

No Discussion

Transmitting Terminal Apparatus for NTSC Color Television

PAGE BURR
NONMEMBER AIEE

DURING the last few years several laboratories in the United States have been carrying on investigations directed toward supplementing the present monochrome television service with transmissions in color. In most cases these investigations of color television systems have been confined to proposals which were expected to receive the wide public acceptance which has characterized monochrome television. The probable commercial success of a color television system may be estimated by noting the degree to which it satisfies certain important criteria. One of these resides in the fact that the system should be "compatible." That is, it is desirable, that television broadcasts in color should be reproduced in monochrome by the 20,000,000 odd black-and-white receivers now in the hands of the public without the necessity of modifying these receivers in any way. Second, in addition to providing full color to a color receiver, the system should provide as high a degree of pictorial definition to both monochrome and color receivers as is now realized in the monochrome service. Third, as a purely technical matter, the radio-frequency spectrum space allocated to the television service should be utilized as efficiently as possible. The color television system which is now being field-tested by the National Television System Committee (NTSC)[1,2] represents the results of these investigations. This system satisfies these criteria in large measure.

At the present writing the details of the NTSC standards are receiving intensive study and test, and it is possible that some alterations of the November 1951 field-test specifications may be adopted. The techniques described in the present paper are necessarily based on the existing field-test specifications. It is believed that the principles and methods used will in general be of continuing interest, although some particulars of the standards may later be different, requiring corresponding modifications of the equipment.

Color Specification in the NTSC System

The NTSC system specifies color values in television images by the use of several concepts which, although common in the science of colorimetry, may not be familiar to workers concerned with other uses of additive or subtractive color systems. It is common knowledge that any color may be specified in terms of the intensities of three primary lights which, when added together, achieve the subjective effect of a "match" with the sample. An important restriction upon the character of these primaries is that they must be independent, that is, no combination of two of them may be capable of matching the third. In practice these lights are usually red, green, and blue.[2–6] As a result, initial investigations in the field of color television were confined to systems wherein the color values of the scene in terms of red, green, and blue were sequentially or simultaneously transmitted to the remote receiver by means of one or

---

Paper **53-91**, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted October 23, 1952; made available for printing December 1, 1952.

PAGE BURR is with Hazeltine Electronics Corporation, Little Neck, N. Y.

The NTSC transmission proposals represent the combined efforts of many workers in various fields of television engineering. The apparatus arrangement described in this paper was originally conceived by Bernard D. Loughlin, and the author is greatly indebted to him for guidance. Thanks are also due to Honer Webb and Frank Wilczek for their assistance during development of the equipment. The preparation of the text, references, and figures for this report has been greatly facilitated by Charles E. Dean; his contribution is most gratefully acknowledged.

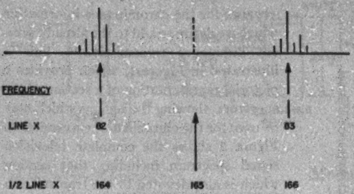


Figure 1 (left). Section of spectrum

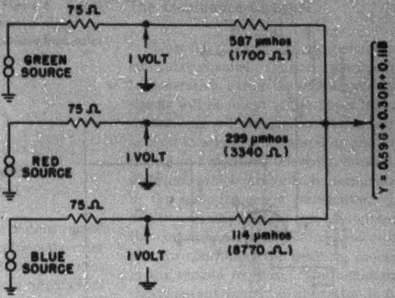


Figure 3 (right). Composition

more separate television channels.[7,8] This practice is strictly analogous to the process of making color separations in color photography, as, for example, in the use of the so-called "1-shot" color camera.[9,10]

As work on various color television systems progressed it became apparent that several alternative methods of color specification were more attractive for color television transmission. One of these alternate methods describes a colored area by means of three quantities proportional to subjective brightness, hue, and saturation. Brightness is closely connected with intensity and is an idea with which we are generally familiar. The concept of hue relates to the nearest shade among the pure spectrum colors or among the purples (that is, mixtures of spectral red and blue) which most closely matches the sample. Saturation is a measure of the degree to which the particular color is pure in the sense of being free of dilution with white or gray light. As an example, pastel or tinted shades are usually characterized by low saturation. In color television terminology the brightness attribute of a colored area is termed its *luminance*. The qualities of the sample relating to colored characteristics, that is, hue and saturation, are termed *chrominance*. In more specific terms chrominance represents the colorimetric difference between any color and a reference gray or white of equal luminance. For the NTSC transmission this reference shade corresponds to average daylight, and is known as Standard Source C of the Commission Internationale d'Eclairage.[11]

A series of experiments fundamental to the development of the NTSC transmission has established that the ability of the eye to resolve fine details of contrasting luminance in a picture is approximately 60 per cent greater than the corresponding ability to resolve fine details which contrast only in chrominance.[12] Consequently, it became apparent at an early stage that the preservation of high definition in the transmission of color television images was very largely controlled by the provisions made for handling the luminance detail.

Relationship to Present Monochrome Signal

The significance of the NTSC color specifications will become apparent when it is noted that present-day monochrome transmissions provide the receiver with information which is approximately representative of the luminance of the scene. As a result, the major problem to be resolved during the development of the NTSC transmission scheme has been to devise a method for incorporating chrominance data into the monochrome service without degrading the performance of existing black-and-white receivers. These receivers would therefore continue to be useful and would serve to reproduce in monochrome the programs transmitted in color.

The method by which this incorporation has been contrived is most easily understood through the application of Fourier series analysis to the video signal wave form of a monochrome transmission. The spectral distribution revealed by this analysis extends to a frequency of approximately 4 megacycles. However, examination of the spectrum indicates that the energy is not continuously distributed, but rather is present only in discrete bundles centered around specific frequencies[13] which are harmonics of the line repetition rate of the television scanning process. A considerable portion of the spectrum of a present-day monochrome transmission is, therefore, in a quite literal sense, not utilized. This statement may be easily verified by the owner of a reasonably selective high-frequency communications receiver. Further, it may be shown that if a signal component (as for example a continuous carrier wave) is introduced into one of these unused "gaps" in the spectrum, then the result would not be perceptible upon an ideal television display. In practical receivers, however, nonlinear phenomena, together with certain properties of the eye, compromise this ideal "invisibility" to a point where the visibility of the signal component is low, rather than zero.

Considerable experimental evidence has accumulated to substantiate the conclusion that these gaps in the television transmission spectrum will accommodate additional signal components without seriously degrading the quality of the received images. It is immediately apparent, therefore, that these unused portions of the spectrum provide a suitable con-

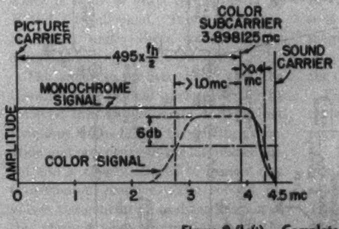


Figure 2 (left). Complete spectrum

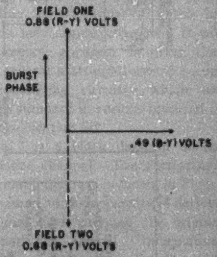


Figure 4 (right). Phase relations of NTSC chrominance subcarrier

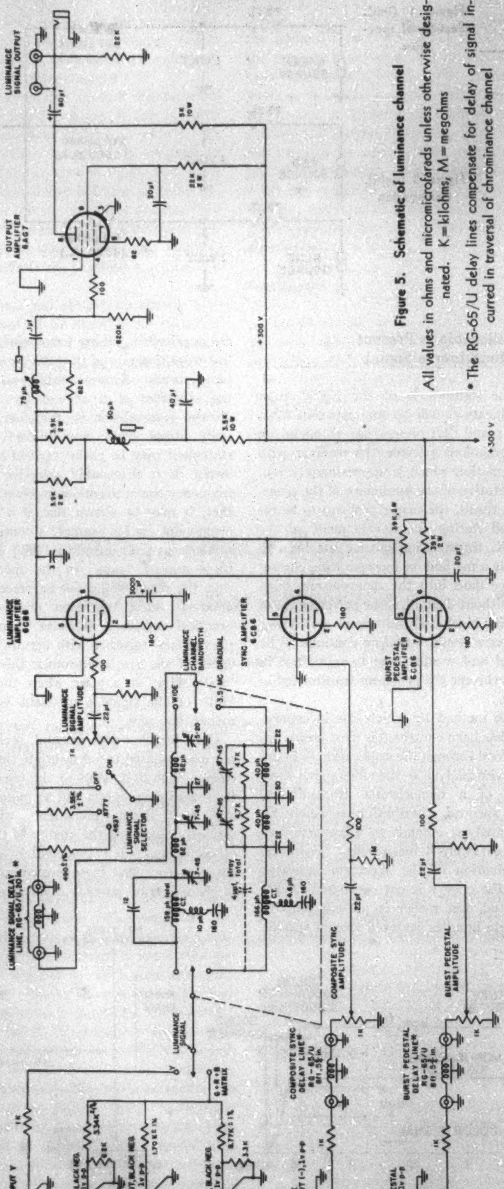

LUMINANCE SIGNAL OUTPUT
OUTPUT AMPLIFIER 6AG7
LUMINANCE AMPLIFIER 6CB6
SYNC AMPLIFIER 6CB6
BURST PEDESTAL AMPLIFIER 6CB6
LUMINANCE SIGNAL AMPLITUDE
LUMINANCE SIGNAL SELECTOR
LUMINANCE CHANNEL BANDWIDTH
WIDE
3.5 MC GRADUAL
LUMINANCE SIGNAL DELAY LINE, RG-65/U, 20 in.*
COMPOSITE SYNC AMPLITUDE
BURST PEDESTAL AMPLITUDE
LUMINANCE SIGNAL
G+R+B MATRIX
LUMINANCE INPUT Y
RED INPUT, BLACK NEG. 1v P-P
GREEN INPUT, BLACK NEG. 1v P-P
BLUE INPUT, BLACK NEG. 1v P-P
COMPOSITE SYNC INPUT (-), 1v P-P
COMPOSITE SYNC DELAY LINE* RG-65/U
BURST PEDESTAL INPUT (-), 1v P-P
BURST PEDESTAL DELAY LINE* RG-65/U
+300 V

Figure 5. Schematic of luminance channel

All values in ohms and micromicrofarads unless otherwise designated. K = kilohms, M = megohms

* The RG-65/U delay lines compensate for delay of signal incurred in traversal of chrominance channel

veyance for the chrominance information which we desire to add to the already present luminance transmission.[14] This is illustrated in Figure 1, which provides a pictorial representation of a section of the spectrum, showing the regions which may be used for the chrominance transmission. Figure 2 shows the complete television signal spectrum including that section which is also devoted to the transmission of chrominance information. In the NTSC transmission the chrominance information is inserted as suppressed-carrier modulation[15—18] of a frequency in the vicinity of 3.89 megacycles. This subcarrier frequency is chosen in such a way that both the subcarrier and its side bands fall in the gaps of the luminance signal spectrum. It will be noted that the chrominance information is shown as occupying an effective band of approximately 1.6 megacycles on a vestigial-sideband basis. This frequency range is sufficient in view of the comparatively low definition required of chrominance information.

Electrical Representation of Luminance and Chrominance

Before proceeding to a general discussion of transmitting terminal apparatus, it will prove advantageous to describe the electrical nature of the luminance and chrominance signals.

As a matter of practical expediency it must be assumed for the present that the pickup device functions in a manner analogous to that of the 1-shot photographic camera referred to previously. That is, the color camera in the television studio will provide the signal-processing apparatus with three video signals which are representative of the green, red, and blue values of the scene. The output signals from such a camera are therefore not directly suited for transmission by a system of the type described here, and must be converted, by a simple co-ordinate transformation, into data of the desired form.

The luminance signal may be obtained by properly weighting the green, red, and blue camera outputs according to their luminance contributions to the reproduced image and then adding them. That is

$$Y = 0.59G + 0.30R + 0.11B \qquad (1)$$

where Y represents the luminance signal voltage, and R, G, and B denote the camera output signal voltages. The three numerical coefficients are determined by the relative luminosities of the three particular reproducing primaries employed at the receiver and by known properties

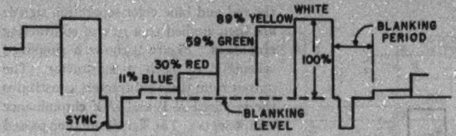


Figure 6. Luminance signal resulting from color bars

of the average human eye.

A method of performing this synthesis of the luminance signal is shown in Figure 3. In this diagram the green, red, and blue signal sources are presumed to be adjusted so that, when the camera is scanning a portion of a scene which is white or gray (no color), the values of the three output voltages are equal. The network of three resistors connected to a common output junction establishes that the summation voltage on this junction will be in proportion to the coefficients of equation 1.

As previously stated, chrominance describes all those attributes of the scene which are not embraced by the luminance information. Also, chrominance requires two quantities for its specification. These two "dimensions" may be obtained electrically by forming the "difference signals,"

$$\begin{array}{l} R-Y \\ B-Y \\ G-Y \end{array} \qquad (2)$$

Any two of these three voltages completely specify the chrominance signal. That is, the quantity $G-Y$ is not independent of $R-Y$ and $B-Y$, but may be expressed in terms of these two chrominance components by substituting equation 1 into equations 2 as follows

$$(G-Y)=a(R-Y)+b(B-Y) \qquad (3)$$

or

$$0.41G-0.30R-0.11B=a(0.70R-0.59G-0.11B)+b(0.89B-0.59G-0.30R) \qquad (4)$$

Hence, since G, R, and B are independent of each other, we have upon equating coefficients for R and B respectively

$$\begin{array}{l} 0.70a-0.30b=-0.30 \\ -0.11a+0.89b=-0.11 \end{array} \qquad (5)$$

Therefore

$$a=-0.51,\ b=-0.19$$

and

$$(G-Y)=-0.51(R-Y)-0.19(B-Y) \qquad (6)$$

It is of interest to note that the original red, green, and blue camera signals can be reconstituted by a proper combination of any two of the chrominance components with the luminance signal.

Modulation of Color Subcarrier by Chrominance Signal

The modulation of this 2-dimensional signal upon the single color subcarrier wave at approximately 3.89 megacycles may be accomplished by a number of means. Fortunately, as is well known, there are two parameters describing a sinusoidal current which may be varied independently. The first of these represents the amplitude of the wave while the second denotes its phase with respect to some reference standard. The chrominance information is modulated upon the subcarrier by means of a technique of this type. It might be said that the amplitude of the modulated subcarrier is a function of the saturation of a colored area while the phase is indicative of the hue.

Although the end result of the modulation process may be rigorously regarded in this light, it will be considerably more convenient to consider that the modulated subcarrier wave is actually composed of two orthogonal sinusoidal components which are independently modulated in amplitude. In particular, Figure 4 provides an illustration of the modulated chrominance subcarrier components. Inasmuch as the vector field represented by the chrominance signal is 2-dimensional, there are a number of axes upon which orthogonal projections of the signal characteristics could be taken. In the example shown the vectors are **R-Y** and **B-Y**. The coefficients of the respective signals indicate their normalized magnitude relative to a luminance signal of unit value. It has been implied previously that for any portion of the televised scene which is white or gray, the values of R, B, and Y are all equal. It is of interest to note that under these conditions the chrominance components vanish.

The paper so far has shown that the transmitting terminal apparatus may be resolved into two separate and distinct functions. The first of these is the generation and processing of the luminance signal and is closely related to present-day transmission practice. The second involves the generation and modulation of the chrominance information and the insertion of the modulated subcarrier into the luminance transmission channel.

The Luminance Channel

Figure 5 is a schematic of the complete luminance channel. The resistive matrix for generating Y from the color-separation signals will be found at the upper left of the drawing. This signal is subsequently applied to a luminance amplifier and delivered to a standard television transmission line, which presumably terminates in the radio-frequency transmitter. In addition an amplifier is provided to insert the conventional synchronizing information for controlling the scanning process in the remote receiver. The luminance signal output from this apparatus is indistinguishable from a conventional black-and-white television video signal as it might exist at various points in the studio facilities of a broadcaster.

Figure 6 shows an example of the video signal which would be generated if the color camera were directed at a geometrical pattern of vertically disposed colored bars. The order of the bars horizontally across the picture is shown as blue, red, green, yellow, and white. The wave form indicates that the luminance contribution of the primary color bars is in direct proportion to the coefficients of equation 1. The luminance contribution from the yellow bar is that due to the sum of red and green. The white bar, being composed of all three primaries, displays a value of 1.0 in luminance. The synchronizing signal and the blanking period are also shown by the diagram. It is perhaps of interest to note that the luminance signal in a color television transmission represents the so-called orthochromatic representation of color values in black and white which is familiar to workers in photography.

The Chrominance Channel

At first sight the apparatus for generating the chrominance information and modulating the subcarrier with this information may appear complex. It will shortly become apparent, however, that it is in general a straightforward equipment embodying circuitry which has long been standard in various branches of the communication art.

Figure 7 is a block diagram of the chrominance channel. The corresponding schematic diagram is shown in Figure 8. The input to the apparatus is derived from signal lines carrying the original color-separation signal information. These color signals are applied to a matrix of the type previously described in order to obtain the luminance component Y. The luminance signal together with

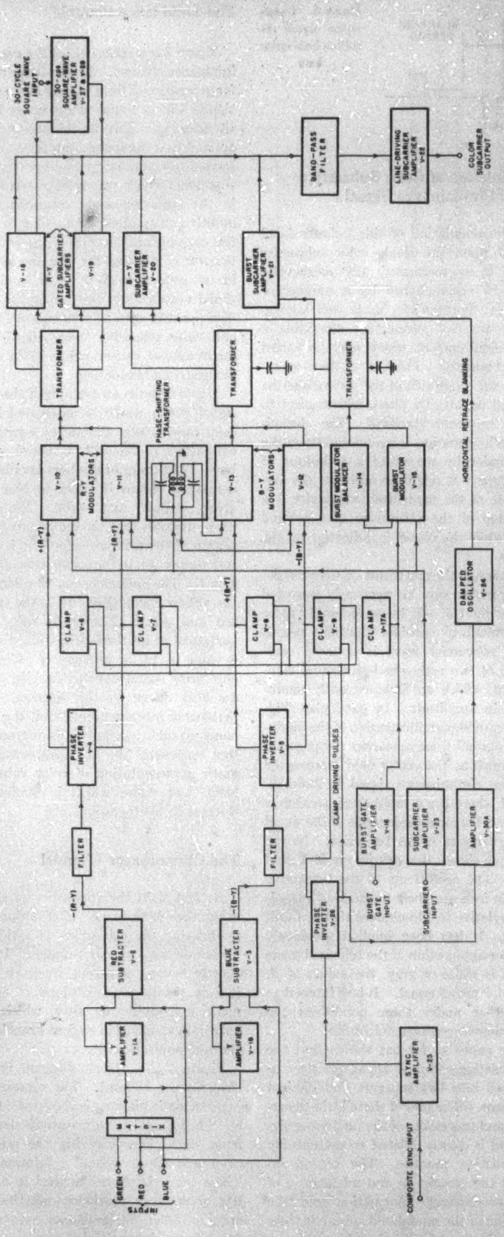
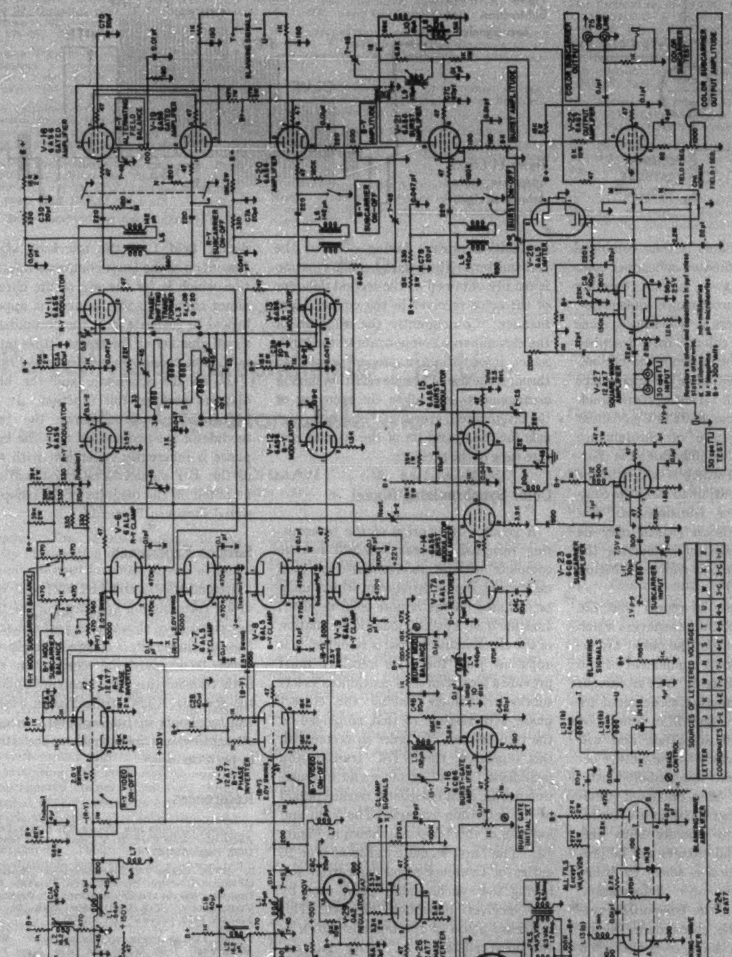

30-CYCLE SQUARE WAVE INPUT
30-CPS SQUARE-WAVE AMPLIFIER V-27 & V-29
BAND-PASS FILTER
LINE-DRIVING SUBCARRIER AMPLIFIER V-22
COLOR SUBCARRIER OUTPUT
V-18
R-Y GATED SUBCARRIER AMPLIFIERS
V-19
B-Y SUBCARRIER AMPLIFIER V-20
BURST SUBCARRIER AMPLIFIER V-21
TRANSFORMER
TRANSFORMER
TRANSFORMER
PHASE-SHIFTING TRANSFORMER
HORIZONTAL RETRACE BLANKING
V-10
R-Y MODULATORS
V-11
V-13
B-Y MODULATORS
V-12
BURST MODULATOR BALANCER V-14
BURST MODULATOR V-15
+(R-Y)
-(R-Y)
+(B-Y)
-(B-Y)
DAMPED OSCILLATOR V-24
CLAMP V-6
CLAMP V-7
CLAMP V-8
CLAMP V-9
CLAMP V-17A
PHASE INVERTER V-4
PHASE INVERTER V-5
CLAMP DRIVING PULSES
BURST GATE AMPLIFIER V-16
SUBCARRIER AMPLIFIER V-23
AMPLIFIER V-30A
BURST GATE INPUT
SUBCARRIER INPUT
FILTER
FILTER
-(R-Y)
-(B-Y)
PHASE INVERTER V-26
RED SUBTRACTER V-2
BLUE SUBTRACTER V-3
Y AMPLIFIER V-1A
Y AMPLIFIER V-1B
SYNC AMPLIFIER V-25
MATRIX
GREEN
RED
BLUE
INPUTS
COMPOSITE SYNC INPUT

Figure 7. Block diagram of chrominance channel

the red and blue color-separation signals are then applied to a pair of subtracting amplifiers. Figure 9 shows a simplified schematic of such a subtracter. The output from these subtracters constitutes the required *R-Y* and *B-Y* chrominance components. The signals are then passed through low-pass filters which limit the chrominance information to a spectrum extending from zero to approximately 2 megacycles. The filter networks are followed by a pair of phase inverters in order to make the chrominance components available in opposing polarities or "push-pull." Conventional video clamping circuits are then employed to establish the potential of black in the signal at a specified value. The video circuit manipulations of the signal are completed at this point.

The next step involves the actual modulation process. To this end two sets of doubly balanced modulators are employed. This arrangement has been extensively used in carrier telephone practice on wire and radio circuits. Both the modulating signal and the local subcarrier source are injected in phase opposition so that the output of the modulator contains neither carrier nor modulation signals but only the modulation side bands. The necessary orthogonal relationship of the chrominance signal vectors is obtained by exciting the *B-Y* modulator with a subcarrier source differing in phase by 90 degrees from that applied to the *R-Y* modulator.

It has been considered desirable up to the present time to reverse the phase sequence of the radiated chrominance subcarrier on alternate fields of the television signal. This practice ameliorates certain difficulties which are inherently characteristic of vestigial-side-band operation. The necessary phase sequence reversal is accomplished by alternately inverting the polarity of the modulated *R-Y* chrominance signal component. This operation is illustrated in the vector diagram, Figure 4. During field 1 of the television transmission the vector diagram is as shown by the solid line, while for field 2 the **R-Y** vector occupies the position indicated by the dotted line. Phase reversal of the *R-Y* component is obtained by alternately selecting one of two outputs from the phase-inverting transformer following the *R-Y* modulators as shown in the block diagram. This selection is performed by a pair of gated amplifiers *V-18* and *V-19*, which are alternately switched into conduction by a 30-cycle square wave which is related to the field being transmitted and which is generated within the system synchronizing apparatus. The output of



V-1A 12AT7 LUMINANCE AMPLIFIER
V-1B 12AT7 LUMINANCE AMPLIFIER
V-2 6CB6 RED SUBTRACTER
V-3 6CB6 BLUE SUBTRACTER
V-4 12AT7 R-Y PHASE INVERTER
V-5 12AT7 B-Y PHASE INVERTER
V-6 6AL5 R-Y CLAMP
V-7 6AL5 R-Y CLAMP
V-8 6AL5 B-Y CLAMP
V-9 6AL5 B-Y CLAMP
V-10 6AS6 R-Y MODULATOR
V-11 6AS6 R-Y MODULATOR
V-12 6AS6 B-Y MODULATOR
V-13 6AS6 B-Y MODULATOR
V-14 6AS6 BURST MODULATOR BALANCER
V-15 6AS6 BURST MODULATOR
V-16 6CB6 BURST GATE AMPLIFIER
V-17A 6AL5 D-C RESTORER
V-18 6AS6 GATED AMPLIFIER
V-19 6AS6 GATED AMPLIFIER
V-20 6AS6 GATED AMPLIFIER
V-21 6AS6 BURST AMPLIFIER
V-22 6AG7 OUTPUT AMPLIFIER
V-23 6CB6 SUBCARRIER AMPLIFIER
V-24 12AT7
V-25 6CB6 SYNC AMPLIFIER
V-26 12AT7 PHASE INVERTER
V-27 12AT7 SQUARE-WAVE AMPLIFIER
V-28 6AL5 LIMITER
V-29 0A2 REGULATOR
V-30A 12AT7 SAWTOOTH GENERATOR
RED INPUT TEST
RED 1V P-P BLACK NEG.
GREEN INPUT TEST
GREEN 1V P-P BLACK NEG.
BLUE INPUT TEST
BLUE 1V P-P BLACK NEG.
HOR. LINE DRIVE TEST
HOR. LINE DRIVE INPUT
BURST GATE INPUT TEST
BURST GATE INPUT 3V P-P NEG.
SUBCARRIER INPUT TEST
SUBCARRIER INPUT 1V P-P
Y INPUT GAIN
INPUT SIGNAL SELECTOR
R-Y VIDEO ON-OFF
B-Y VIDEO ON-OFF
R-Y MOD. SUBCARRIER BALANCE
B-Y MOD. SUBCARRIER BALANCE
BURST MOD. BALANCE
BURST GATE INITIAL SET
R-Y SUBCARRIER ON-OFF
B-Y SUBCARRIER ON-OFF
B-Y AMPLITUDE
BURST ON-OFF
BURST AMPLITUDE
R-Y ALTERNATING FIELD BALANCE
PHASE-SHIFTING NETWORK
BLANKING SIGNALS
CLAMP SIGNALS
BIAS CONTROL
BLANKING-WAVE AMPLIFIER
BLANKING-WAVE SHAPER
SUBCARRIER INPUT
30 cps TU INPUT
30 cps TU TEST
COLOR SUBCARRIER OUTPUT
COLOR SUBCARRIER TEST
COLOR SUBCARRIER OUTPUT AMPLITUDE
+133V
+150V
SOURCES OF LETTERED VOLTAGES
LETTER | J | K | M | N | S | T | U | W | X | Z
COORDINATES | 5-E | 4-E | 7-A | 7-A | 4-E | 4-A | 4-A | 3-C | 3-C | 1-A
A
B
C
D
E
1
2
3
4
5
6
7
8

Figure 8. Schematic of chrominance channel of encoder

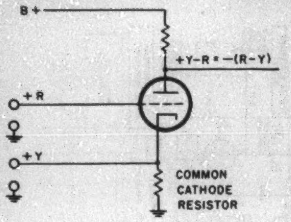


Figure 9 (left). Subtraction of two signals

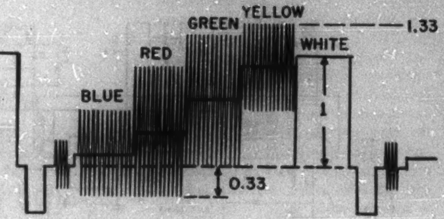


Figure 11 (right). Complete color television signal

these gated amplifiers is then combined with the *B-Y* modulated subcarrier component. These two signals together comprise the chrominance subcarrier. If the studio camera is televising a scene composed of the colored bars described previously, the output from the chrominance channel would have the appearance shown in Figure 10. That is, corresponding to each color bar there is a particular amplitude and phase of chrominance signal. The diagram illustrates the character of the signal which is added to the luminance signal output to form the complete NTSC color transmission. The result of this addition (which may be accomplished simply by connecting the two cables together) is shown in Figure 11.

It is worth while to repeat that the conventional monochrome receivers which are now in the hands of the public are responsive only to that component of the signal shown in Figure 6. The additional chrominance information as detected by a monochrome receiver will in general be low in reproduced amplitude and will cause little degradation to the picture because of the inherent low-visibility properties which have been discussed previously. On the other hand, in a color receiver designed to operate with this transmission, special circuits will be provided whose function is solely to utilize the added chrominance information. The magnitudes of the component vectors composing the chrominance signal will be detected in the receiver through a process of synchronous demodulation which is essentially the reverse of the modulation process taking place at the transmitter. The luminance information is readily obtained at the second detector of the color receiver in the conventional manner. Consequently the addition of the chrominance components to the luminance information in proper proportion through the use of simple resistive mixing arrangements will allow the recovery of the original red, green, and blue color signals for the operation of the tricolor picture tube of the receiver.

Color Synchronizing Signal

A general description of the transmitting terminal apparatus would not be complete without reference to the synchronizing signal provisions which must be included for the remote color receiver. Inasmuch as synchronous demodulation of the chrominance signal is required, it is apparent that the color receiver must provide a local source of continuous color-subcarrier voltage having the correct phase with respect to that employed at the transmitter. In order to accomplish this objective, the NTSC transmission includes a "burst" of subcarrier information during each horizontal retrace period of the television system. The form and location of this burst are shown in Figure 12. The burst signal consists of 9 or 10 cycles of reference-phase subcarrier following the synchronizing pulse used by the receiver deflection circuits. This reference phase of the burst is shown as the burst phase in the vector diagram, Figure 4, and coincides with that of the *R-Y* subcarrier component transmitted during field 1 of the television signal. Inasmuch as the burst comprises information which is for the use of the chrominance circuits in the receiver, it appears logical to generate it in the chrominance signal channel at the transmitting terminal. The tubes labelled *V-14*, *V-15*, and *V-21* on the schematic and the block diagram perform this function. In the particular apparatus shown the burst modulator is so connected that the burst phase is inherently coincident with that of the *R-Y* signal. The burst will also be noted on the diagram of the complete signal shown in Figure 11.

Service Experience

Extended laboratory trials of color television transmission apparatus of the type described have shown it to perform satisfactorily in service. There is ample evidence to support the expectation that both transmitting and receiving apparatus for the NTSC transmission system will be simple in fundamental concepts, straightforward in execution, and stable in performance.

References

1. Specifications for Color TV Field Tests. *Electronics* (New York, N. Y.), volume 25, January 1952, pages 126–27.

2. NTSC Color-TV Specifications: Summary of Standards to Be Used in Testing All-Industry Compatible System in Principal Cities Soon. Tele-Tech (New York, N. Y.), January 1952, pages 61 and 102–03.

3. An Introduction to Color (book), Ralph M. Evans. John Wiley and Sons, New York, N. Y., 1948.

4. Color in Business, Science, and Industry (book), Deane B. Judd. John Wiley and Sons, New York, N. Y., 1952.

5. Color Fundamentals for TV Engineers,

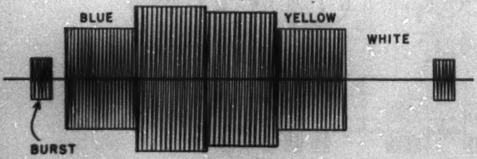


Figure 10. Chrominance signal corresponding to color bars

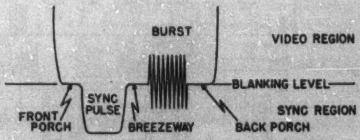


Figure 12. Horizontal retrace interval, showing burst

Donald G. Fink. *Electronics* (New York, N. Y.), volume 23, December 1950, pages 88–93; volume 24, January 1951, pages 78–83; February 1951, pages 104–09; March 1951, pages 336 and 338.

6. COLOR TELEVISION AND COLORIMETRY, W. T. Wintringham. *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 39, October 1951, pages 1135–72.

7. AN EXPERIMENTAL SIMULTANEOUS COLOR TELEVISION SYSTEM, R. D. Kell et al. *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 35, September 1947, pages 861–75.

8. COLOR TELEVISION—U.S.A. STANDARD, Peter C. Goldmark. *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 39, October 1951, pages 1288–1313.

9. A HALF CENTURY OF COLOR (book), L. W. Sipley. The Macmillan Company, New York, N. Y., 1951, pages 58–62 and 86–109.

10. PHOTOGRAPHY: ITS MATERIALS AND PROCESSES, C. B. Neblette. D. Van Nostrand Company, New York, N. Y., fifth edition, 1952, pages 408–11.

11. THE MEASUREMENT OF COLOUR (book), W. D. Wright. Adam Hilger, London, England, 1944, pages 204–06.

12. MIXED HIGHS IN COLOR TELEVISION, A. V. Bedford. *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 38, September 1950, pages 1003–09.

13. TELEVISION—THE SCANNING PROCESS, Pierre Mertz. *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 29, October 1941, pages 529–37.

14. PRINCIPLES OF NTSC COMPATIBLE COLOR TELEVISION, C. J. Hirsch, W. F. Bailey, B. D. Loughlin. *Electronics* (New York, N. Y.), volume 25, February 1952, pages 88–95.

15. DUPLEX TRANSLATING CIRCUITS, J. R. Carson. United States Patents *1,343,306* and *1,343,307*, June 15, 1920.

16. METHOD AND MEANS FOR SIGNALLING WITH HIGH FREQUENCY WAVES, J. R. Carson. United States Patent *1,349,382*, March 27, 1923.

17. PRODUCTION OF SINGLE SIDEBAND FOR TRANS-ATLANTIC RADIO TELEPHONY, R. A. Heising. *Proceedings*, Institute of Radio Engineers, (New York, N. Y.) volume 13, June 1925.

18. RADIO ENGINEERS' HANDBOOK (book), F. E. Terman. McGraw-Hill Book Company, Inc. (New York, N. Y.), 1943, pages 551–53.

No Discussion

Polyethylene Terephthalate—Its Use as a Capacitor Dielectric

M. C. WOOLEY
NONMEMBER AIEE

G. T. KOHMAN
NONMEMBER AIEE

W. McMAHON
NONMEMBER AIEE

Synopsis: The steady increase in the severity of operating conditions for capacitors and the need for more diverse characteristics has spurred the search for new and better capacitor dielectrics. The synthetic plastics industry, which is the source of a number of useful dielectric materials, has recently produced a new and promising dielectric in film form known as polyethylene terephthalate or "Mylar." This material is unusually strong, has a high softening point, and is available in very thin films which makes it especially suitable for capacitor insulation. The electrical characteristics of capacitors wound with Mylar film are likewise promising. As compared to mineral-oil-impregnated paper capacitors, unimpregnated Mylar capacitors exhibit higher dielectric strength, higher insulation resistance, and can be operated at higher ambient temperatures. Their loss characteristics are comparable with those of impregnated paper and the capacitance stability over the usual range of ambient temperatures approaches that of mica capacitors. Mylar is relatively nonhygroscopic and tests indicate that for moderate atmospheric conditions capacitors made from it do not require additional moisture protection. The film can be metallized readily by current techniques and, when used in metallized capacitors, appears to possess advantages over metallized paper in several respects.

The Need for New Dielectrics

THE diversity of requirements for electrical insulating materials is seemingly ever increasing and the urge to meet these demands is great. This has caused a continual scanning of new materials with the hope of finding some with better properties or a more advantageous combination of properties than those now in use. It is to be expected that the chance of finding such a material is small because the insulating materials now in use have been selected with great care and many improvements have been made in them. For example, paper has been used as insulation in wound capacitors for many years. It is an excellent material for this purpose and difficult to surpass. Nevertheless, it does have some inherent shortcomings. Paper decomposes at elevated temperatures and the trend in recent years toward higher operating temperatures for electric equipment has emphasized this restriction on its use. Also paper is hygroscopic and this necessitates thorough protection against atmospheric moisture which adds greatly to the cost and size of capacitors. In addition, the quality of paper is rather difficult to control because the raw material, wood, or other cellulosic plant fiber, from which paper is made is variable in composition; and too, the paper-making process is quite complex. Thus, one sees that there is ample incentive to replace or supplement even so firmly entrenched a material as paper.

General Properties

One of the more fruitful fields in the search for better materials is the synthetic plastics industry. A number of insulating materials for a variety of purposes have come from this field and recently a very promising one, polyethylene terephthalate, has been added to the list. This material, variously known as Terylene, *V-200* film, or Mylar film, promises to be outstanding for capacitor insulation and may also be adaptable to coil interleaving. Terylene is the name used by Imperial Chemicals, Ltd., of Great Britain, while the DuPont Company which developed manufacture of the material in this country called it *V-200* film during the early stages of development but have now adopted the trade name, Mylar, for the film.

Mylar is a polyester of ethylene glycol and terephthalic acid. It is a colorless transparent material with a specific gravity of about 1.4. It is exceptionally tough, having a high tensile strength and much better tear resistance than most plastic films. For this reason it is feasible to make very thin films which is a decided advantage for capacitor insulation. The softening point of the film

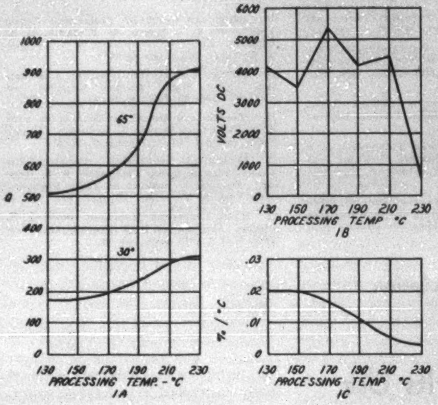
Table I. Dielectric Properties of Mylar

Frequency, Kc	Dielectric Constant at 25 Degrees Centigrade	Power Factor at 25 Degrees Centigrade
1	3.4	0.0055
10	3.4	0.018
100	3.3	0.016

D-c resistivity at 25 degrees centigrade (400 volts) 3×10^{15} ohm-centimeters
Dielectric breakdown strength at 25 degrees centigrade (single layer of 1.33-mil film)—6.0 kv/mil
Conducting paths in 1.33-mil film—nil

Paper 53-6, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted April 11, 1952; made available for printing October 7, 1952.

M. C. WOOLEY, G. T. KOHMAN, and W. MCMAHON are with Bell Telephone Laboratories, Inc., Murray Hill, N. J.



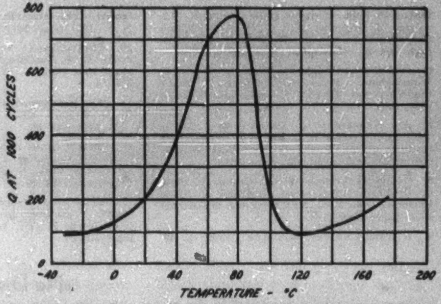


Figure 1 (left). The effects of capacitor unit processing temperature on (A) Q at 1,000 cycles, (B) ultimate dielectric strength, and (C) temperature coefficient of capacitance in per cent per degree centigrade

Figure 2 (above). The effect of temperature on Q at 1,000 cycles of a Mylar capacitor

is about 240 degrees centigrade. However, under proper conditions it can be fused at somewhat lower temperatures. Mylar film is quite stable chemically. It is highly resistant to moisture penetration and also to decomposition by hydrolysis. Since chemical and physical properties such as these are admirably suited to electrical insulation, the dielectric properties of the material were investigated. The results of preliminary work are shown in Table I. The dielectric breakdown strength of the film is found to be approximately 50 per cent greater than that of impregnated paper. At room temperature the d-c resistivity of Mylar is very high, its power factor is about the same as that of impregnated paper, and its dielectric constant is intermediate between paper and the hydrocarbon polymers, polysytrene, or polyethylene.

Characteristics of Mylar Capacitors

These properties of Mylar film, particularly its high dielectric strength and mechanical strength, high temperature stability and low d-c conductivity, make it especially attractive to the capacitor design engineer. In addition, some of its electrical properties can be improved by suitable heat treatment. This is illustrated in Figure 1 which shows the effect of heat treatment at temperatures up to the softening point of the film. A temperature of approximately 200 degrees centigrade for 2 hours is sufficient to produce the greater part of the benefits of such a treatment but the period of heating is not critical and longer times can be used to obtain some additional improvement. However heating for very much longer time or at much higher temperatures may produce a degradation, especially in breakdown voltage, as illustrated by the last point of the dielectric strength curve in Figure 1. Excessive heating leads to rapid crystallization of the film and the reduction in breakdown voltage of the units is believed to be due to the resulting embrittlement and mechanical damage to the film although there may be a change in its inherent dielectric strength.

DIELECTRIC STRENGTH

The data in Figure 1 were obtained with 0.2-microfarad capacitors having a single layer of Mylar approximately 0.0009 inch thick between electrodes. From the breakdown voltage data it can be seen that an ultimate dielectric strength of 4,000 to 6,000 volts per mil is realized. Tests with thinner single sheets between electrodes have given results averaging as low as 2,000 volts per mil, but it should be emphasized here that these are average results on a number of capacitors each having a total dielectric area of the order of 2 square feet. The usual dielectric strength test using electrodes of small area normally gives values much in excess of those obtained in capacitor units and it would therefore be expected that such tests would give values higher than the maximum given here.

CAPACITANCE STABILITY

Capacitor units which have been heat-treated at temperatures of approximately 200 degrees centigrade have the layers of film fused together and to a considerable degree the film adheres to the foil so that the winding becomes a solid mass. Consequently, such units show a high degree of capacitance stability even when subjected repeatedly to temperatures as high as 150 degrees centigrade.

DIELECTRIC LOSSES

The Q of Mylar (where $Q=(2\pi fC)/G\approx 1/\text{power factor}$) at 1,000 cycles exhibits a maximum at about 80 degrees centigrade as illustrated in Figure 2. At lower frequency the peak is reduced and tends to shift slightly toward lower temperatures while at much higher frequency the curve is relatively flat. The variation in Q at 1,000 cycles with temperature makes direct comparison with other dielectrics difficult but in general it may be said that the Q of Mylar is at least equivalent to that of mineral-oil-impregnated paper over the temperature range of −40 to 100 degrees centigrade. It is inferior at room temperature and below to some impregnated paper capacitors which are designed especially to have low a-c losses. These and the other data reported herein are based on relatively few lots of the film and are not necessarily representative of what may be expected in large scale use of the material. However, there is relatively good agreement between data reported herein and that published by Reddish.[1]

TEMPERATURE COEFFICIENT OF CAPACITANCE

The capacitance-temperature characteristic of Mylar capacitors is shown in Figure 3. Large variations in capacitance are noted but mainly outside the usual range of temperatures. Within the operating range usually encountered, the temperature coefficient of capacitance of suitably processed units is lower than for any of the commonly used dielectric materials excepting mica. Frequency affects the shape of the capacitance temperature curve, and at frequencies higher than the 1,000 cycles for which Figure 3 is plotted the slope of the curve becomes somewhat greater over its central portion.

INSULATION RESISTANCE

Another of the outstanding properties of Mylar is its high resistivity over a very wide temperature range. Of the commonly used capacitor dielectrics only polystyrene exhibits a higher resistivity, and its use has been limited by a relatively low softening point, high cost of thin films, and other factors. While Mylar does not have the extremely high resistivity of polystyrene, its resistivity is much higher than that of other commercially used dielectrics as illustrated in Figure 4 which also shows for comparison the leakage resistance characteristic of a mineral-oil-impregnated paper capacitor. It may be noted that the Mylar capacitor has approximately 50 times the insulation resistance of the paper capacitor over the whole temperature range.

LIFE ON ALTERNATING VOLTAGE

When operated with appreciable amounts of internal a-c corona Mylar film capacitors fail within hours or days, indicating only moderate corona resistance. Unimpregnated Mylar capacitors exhibit corona at 300 to 400 volts rms a-c and impregnation with oil or other suitable material is necessary if alternating potentials greater than these are involved. It is encouraging to note, however, that capacitors do not fail immediately on exposure to severe corona so that premature failure is not likely to result from infrequent short periods of overvoltage.

LIFE ON DIRECT VOLTAGE

The stability of Mylar under high electrical stress and at high temperature is likely to prove one of the most important of its advantages over other capacitor dielectrics. Impregnated paper capacitors are generally limited to operation at temperatures of 100 degrees centigrade or less but preliminary tests indicate that Mylar film capacitors may be used at temperatures at least as high as 150 degrees centigrade. Life tests in which capacitors are operated at a stress of approximately 700 volts per mil at 85 degrees centigrade are usually expected to produce capacitor failures within a period of 1,000 hours or less. In a test with Mylar capacitors at a dielectric stress of approximately 600 volts per mil there have been no failures after 5,000 hours at 130 degrees centigrade plus more than 4,000 hours at 150 degrees centigrade. Much more study will be required before temperature and voltage ratings can be firmly established but these preliminary results are most promising, particularly in the light of experience with other organic dielectric materials.

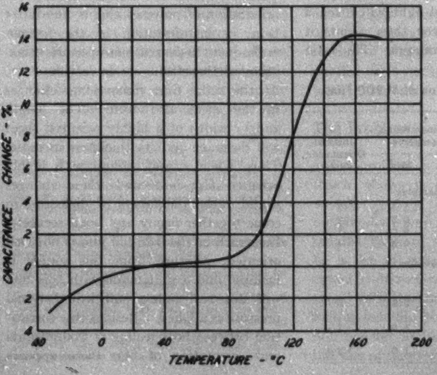
MOISTURE RESISTANCE

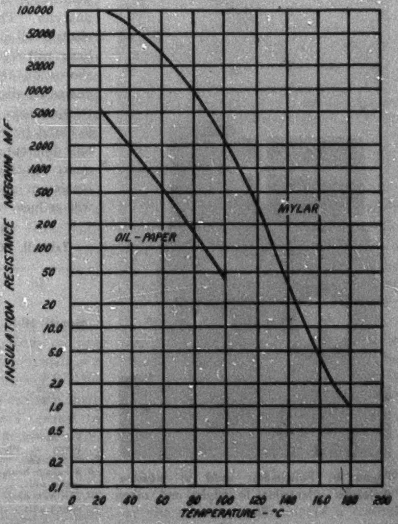
All of the results discussed so far have been obtained with conventionally wound capacitors containing one or two layers of film between aluminum foils and heat-treated after being wound. The use of an impregnant may be necessary under conditions such as previously noted for high alternating voltages but in general it is felt that impregnation is unnecessary or even undesirable. Especially for those applications where the ultimate in insulation resistance or minimum a-c losses are required, the use of dry, heat-sealed Mylar appears to be most suitable. On the other hand, where the highest electrical characteristics are not required but where cost is a primary consideration, Mylar also appears to fill the need. Mylar capacitors are much more inert to the harmful effects of atmospheric moisture than impregnated paper types so that for applications involving moderate humidity conditions we believe that it is possible to use heat-treated Mylar units without additional protection.

An experimental capacitor intended for such application is shown in Figure 5. The terminals of this capacitor are headed wires embedded in the ends of the unit during heat treatment so that they are securely anchored. Heat-treated units of Mylar have withstood continuous

Figure 3 (below). The capacitance-temperature characteristic of a Mylar capacitor at 1,000 cycles

Figure 4 (right). The insulation resistance of a Mylar capacitor is approximately 50 times that of a good mineral-oil-impregnated paper capacitor





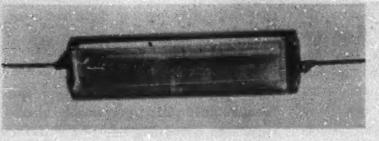
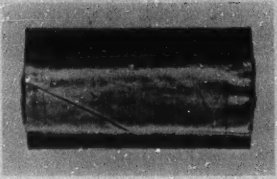


Figure 5 (left). An experimental Mylar capacitor intended to be mounted by its leads

Figure 7 (right). A metallized Mylar capacitor unit



exposure to atmospheric conditions of 90-per-cent relative humidity at 85 degrees Fahrenheit for 7 months with no significant change in insulation resistance. In some instances, microscopic holes along terminals or at other points destroy the effectiveness of the Mylar seal, but vacuum impregnation of the heat-treated capacitor with wax seals those which might otherwise permit moisture to enter. Life tests on continuous voltage of unencased Mylar capacitors such as described also indicate that so long as humidity and temperature conditions are not extreme the life is not significantly affected by exposure to atmospheric conditions.

Film Structure

In heat-sealing capacitors made with films of several different thicknesses, it was found that all except one sealed satisfactorily. Units made with 0.25-mil film did not seal well; that is, in these units the dielectric extending beyond the metal foil electrodes did not fuse together to form a solid transparent mass like the other films. Investigation of this odd behavior led to an interesting correlation between certain physical characteristics of the sheets and heat-sealing properties and also dielectric properties. Despite the fact that wound capacitor units did not seal satisfactorily, sheets of the 0.25-mil grade fused together as well as the other grades when firmly pressed together and heated to the temperature employed to seal the units. This suggested a physical difference in the make-up of the various grades and hence the degree of orientation of polymer chains in the plane of the films was measured.

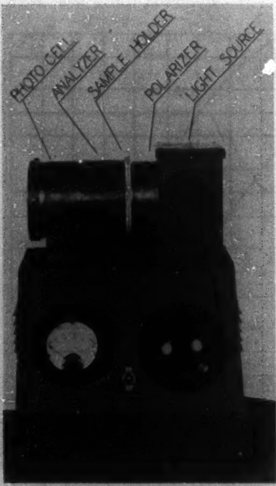


Figure 6. Polarimeter used to measure molecular orientation in Mylar and other films

These measurements were made in the apparatus shown in Figure 6. This is a polarimeter designed to measure orientation in sheets of anisotropic materials. A sample of the film is placed in a rotatable holder between polarizers whose planes of polarization are at right angles to each other. If the film is optically anisotropic, it will twist the plane of the beam from the polarizer so that it may pass through the analyzer. The light falling on the photoelectric cell on the other side of the analyzer will register on the microammeter shown on the panel. If the polymer chains are aligned or oriented with respect to some axis in the plane of the film, it is found by rotating the sample that at four different angular positions 90 degrees apart the transmitted light will reach a maximum and between these points it will reduce to a minimum. The ratio of the maximum to the minimum is a measure of the degree of orientation with respect to a given axis in the plane of the film. This, together with the level of light transmission, serves to indicate the sort of orientation existent in the film. If during rotation of sample the amount of light transmitted changes from very high to very low, uniaxial orientation is indicated. If the level is high for all angular positions of the film, this indicates a planar orientation which might be obtained by simultaneously drawing the film in two directions at right angles to each other. If, however, the light transmission is low at all angular positions, this indicates the absence of orientation at least in the plane of the film. The data in Table II indicate that the 1.0-mil and 0.5-mil films are highly oriented uniaxially, while the comparatively high transmission of 0.35-mil film in all positions indicates some sort of planar or biaxial orientation. The 0.25-mil film, however, shows very little orientation in the plane of the film. Heating the films to temperatures of the order of 400 degrees Fahrenheit results in a considerable amount of deorientation in the films which were highly oriented originally but causes little change in the 0.25-mil film.

These data were interpreted as follows. Extrusion and drawing of the film produces strains which are accompanied by orientation of polymer chains, the latter being a manifestation of the former. Subsequent heat treatment permits relaxation of the strains and deorientation. At the same time dimensional changes in the sheet also occur. The length and/or width of a highly oriented sheet will decrease and its thickness increase. Thus, when a unit wound with highly oriented film is heated, these changes would cause the layers of dielectric to come together firmly and bond securely. Inasmuch as the 0.25-mil film is not well oriented, pressure from this source is lacking, and a unit made with this film will not seal unless sufficient external pressure is applied. Besides this correlation between heat-sealing of wound units and orientation of film there appears

Table II. Orientation of V-200 Film

Sample, Mils	Light Transmission, Microamperes: Max.	Light Transmission, Microamperes: Min.	Uniaxial Orientation Factor
Original Samples			
1.0	25.1	1.5	17
0.5	23.3	0.8	29
0.35	9.0	8.0	1.1
0.25	1.6	0.4	4
Heated Samples			
1.0 —heated 2 hours at 400 deg. Fahr.	16.8	2.9	6
0.5 —from heat-sealed unit	3.0	1.6	2
0.25—from heat-sealed unit	2.5	1.0	2.5

to be some relation between dielectric properties and polymer orientation. It will be recalled that Figure 1 shows a progressive improvement in most properties with increasing temperature of heat treatment. The film used in these tests was highly oriented originally, and it has been shown that heat treatment causes deorientation. This may indicate that the dielectric properties of highly oriented film are inferior to those of film having a low degree of orientation but it is probable that changes in other properties, such as crystallinity, also take place on heating and these may also affect the electrical characteristics of the film. Nevertheless, the data at hand indicate that from the heat-sealing standpoint alone highly oriented film is to be preferred because of the ease with which it may be heat-sealed.

Metallized Mylar

The high mechanical strength and tear resistance as well as the high softening temperature of Mylar film makes it particularly adaptable to current metallizing techniques wherein the metallic film is deposited on the dielectric by condensation of the metal vapor under vacuum.[2,3] Metallized Mylar capacitors such as illustrated in Figure 7 have all the advantages of metallized paper types such as small size and self-healing characteristics and they have the further advantage over paper in that the Mylar film does not need to be lacquered before it is metallized.

Results with metallized Mylar are as yet rather scattered but there is reason to expect that for equivalent performance capacitors made with metallized Mylar will be smaller than those made with metallized paper.

It now appears that Mylar film may become the chief contender for the place of paper as the dielectric in electrical capacitors and further that Mylar capacitors will be superior in several respects, including insulation resistance, temperature coefficient of capacitance, and operating temperature range.

References

1. The Dielectric Properties of Polyethylene Terephthalate, **Wilson Reddish.** *Transactions*, Faraday Society (London, England) volume 46, 1950, page 459.

2. Metallized Paper for Capacitors, **D. A. McLean.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 38, 1950, page 1010.

3. Metallized Paper Capacitors, **J. R. Weeks.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 38, 1950, page 1015.

No Discussion

Digital-to-Analogue Shaft-Position Transducers

S. J. O'NEIL
ASSOCIATE MEMBER AIEE

IN THE design of digital computers the input and output data are often fed in or taken out in analogue form. These analogue forms may be voltages, pressures, meter readings, graphs, and so forth. At some point the output signals usually have to be converted from electronic form to mechanical motions or to indications of some nature.

For some applications it may be necessary to convert the output from digital form to an analogue output in the form of a shaft position. Where the output is displayed visually on a dial, this is the case. If the output must be fed into some analogue device, the analogue input is often a shaft position.

Since the rotation of a shaft is a relatively slow process limited by the moment of inertia of the rotating member, the friction at the bearings or other points of sliding contact, and the torque available to produce the rotation, it is usually necessary to introduce a parallel digit storage register from which the proper analogue outputs are derived. The derivation of a shaft position from such a parallel digit storage register is the subject of this paper.

The Problem

The problem is illustrated by Figure 1 which shows a parallel bank of five switches feeding the input terminals of a 5-digit shaft-position transducer. This parallel bank of switches actually may be a set of relay contacts opened or closed by driving-amplifier stages to correspond with the state of a parallel bank of high-speed flip-flop tubes. The flip-flop register in turn would probably be but one of many such high-speed output registers. Table I shows the method of shaft-position coding assumed. Straight binary coding only is assumed.

A convenient formula for readily determining the angle associated with any binary number is

$$\theta_q = a\left(\frac{360}{2}\right)+b\left(\frac{360}{2^2}\right)+c\left(\frac{360}{2^3}\right)+\ldots$$

where θ_q is the angle associated with the binary number

$a\,b\,c\ldots$

where

$a = 0$ or 1 is the most significant digit.
$b = 0$ or 1 is the second most significant digit.
$c = 0$ or 1 is the third most significant digit, and so forth.

Principles of Operation

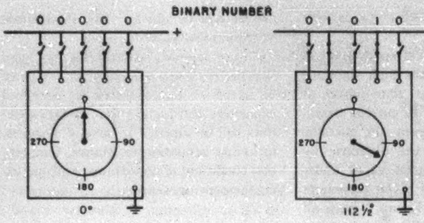
The devices described in this paper depend on the generation of a magnetic field by the stator windings of a synchro or resolver type of output member. The magnetic field can be either a direct or alternating field depending on whether direct voltages or single-phase alternating voltages are applied to the stator windings. The rotor winding can be energized with the corresponding type of voltage and the resulting magnetic field interaction will cause the rotor to turn through the minor arc until a stable relationship between rotor and stator magnetic fields exists. Alternatively, in the a-c case the rotor winding may be used as the secondary winding of a control

---

Paper 53-7, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, January 19–23, 1953. Manuscript submitted July 9, 1952; made available for printing October 8, 1952.

S. J. O'Neil is with the Air Force Cambridge Research Center, Cambridge, Mass.

The author wishes to express his appreciation for the assistance of the members of the Air Force Cambridge Research Center Data-Link Utilization Unit who designed, built, and tested these shaft-position transducers and in particular to Second Lieutenant Robert P. Bigliano, United States Air Force, who invented the digital voltage divider network shown in Figure 6.



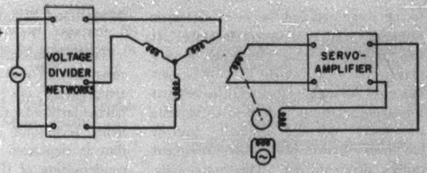


Figure 1 (left). The digital-to-analogue shaft position transducer

Figure 3 (above). Three-phase wound a-c control transformer

transformer which detects the amount and direction of angular rotation for a servomechanism follow-up system.

Figure 2 shows diagrammatically a 2-phase stator type of transducer which has two stator windings 90 degrees apart in space and a single rotor coil with a direct voltage applied to it.

Figure 3 shows a synchro-type transducer which is used as a control transformer to detect an error angle which is nulled by a servomechanism follow-up system.

Magnetic Field Generation

It is well known that the application of three alternating voltages 120 degrees apart in time phase to three windings symmetrically situated on a dynamo stator with their resultant coil axes 120 degrees apart in space phase will cause a rotating magnetic field to be generated in the stator, see Figure 4.

The direction of the instantaneous resultant magnetomotive force at time t_1 is given by the sum of three magnetomotive force vectors 120 degrees apart in space

$$F_r = F_1 + F_2\epsilon^{j120°} + F_3\epsilon^{-j120°}$$

where

Fr = the resultant magnetomotive force at time t_1

F_1 = the magnetomotive force in stator coil number 1 at time t_1

F_2 = the magnetomotive force in stator coil number 2 at time t_1

F_3 = the magnetomotive force in stator coil number 3 at time t_1

Neglecting saturation effects the direction of the resultant magnetomotive force may also be obtained by considering the instantaneous voltages in the stator coils. Since the coil voltages can be related to the applied voltages by means of a well-known circle diagram, the direction of the resultant field can be obtained directly from the applied voltages. If the direction of the instantaneous resultant magnetomotive force is known as a function of the instantaneous values of the applied 3-phase voltages, the angular direction of this field can be substituted as abscissa in Figure 4 in place of time. Therefore, Figure 4 not only can be used to obtain the instantaneous direction of a rotating magnetomotive force at some time t_1 but also can be used to obtain the resultant direction of the magnetomotive force set up by the application of direct voltages of suitable magnitude and polarity to the stator terminals. In other words, three direct voltages related in magnitude and polarity as shown for some angle in Figure 4 and applied to the coils will produce a resultant magnetomotive force in the direction of this angle.

If three single-phase alternating voltages related in polarity and magnitude as shown in Figure 4 for some value of resultant field direction are applied to the stator terminals, a stationary alternating magnetomotive force will be induced in this direction. These are the types of voltages which are normally induced at the stator terminals of a synchro transmitter whose rotor is held at some arbitrary position. In the circuit of Figure 3, for example, the output of the voltage divider network will be the equivalent output of a synchro transmitter.

The same principle can be applied to a machine with two stator windings 90 degrees apart (in space quadrature). The figure corresponding to Figure 4 for this case is Figure 5.

In this case the resultant magnetomotive force is given by

$$F_r = F_1 + F_2\epsilon^{j90°}$$

Since the applied voltages are the coil voltages in this case, the direction of the resultant stator magnetomotive force can be related by a similar equation to the applied voltages.

Linear Voltage Approximations

The sinusoidal voltage amplitude curves shown in Figures 4 and 5 can be followed exactly to generate a very precise angle for a given digital input. Practically, however, this results in the use of a great number of switching elements in the voltage divider networks shown in Figures 2 and 3.

The use of digital information results in a quantization of the angle. For instance, if eight digits of a binary system are used to represent 360 degrees, only $2^8 = 256$ actual angles can be represented. Each angle will, therefore, indicate a range of angles

$$\theta = \theta_q \pm \frac{1}{2}\left(\frac{360}{2^n}\right)$$

where

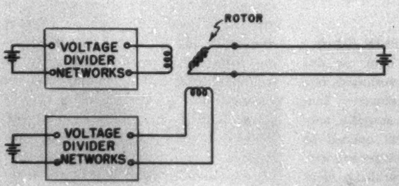


Figure 2. Two-phase wound d-c transducer

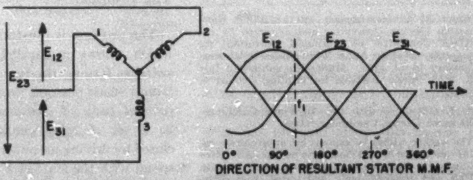


Figure 4. Instantaneous direction of resultant magnetomotive force in a 3-phase wound stator

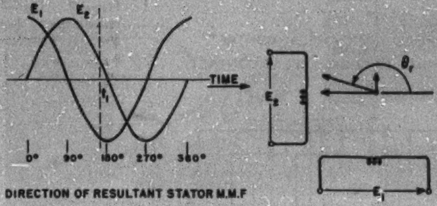


Figure 5 (above). Instantaneous direction of resultant magnetomotive force in a 2-phase-wound stator

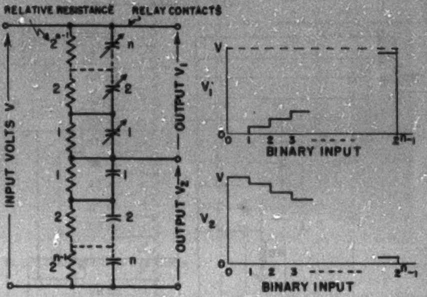


Figure 6 (right). Digital-to-analogue linear voltage divider

θ_q = the quantized angle derived from digital data.
θ = the actual range of angles corresponding to θ_q.
n = the number of digits used to code the angle.

The degree of precision of any angle derived from digital data is $\pm 360/2^{n+1}$ degrees. Therefore, for an efficient overall design the shaft-position transducer need not yield an angle to an accuracy greater than $\pm 360/2^{n+1}$ degrees.

To minimize the number of switching components it is desirable to approximate the sinusoidal voltage amplitude distributions by straight-line segments. A digital-to-analogue linear voltage divider involves fewer components than a sinusoidal voltage divider since linear voltages can be subdivided into binary form and added in combinations to produce any output. A circuit of this type is shown in Figure 6.

The circuit of Figure 6 produces two different discrete voltage outputs whose sum is equal to the input voltage V. The output voltage V_1 is a linearly increasing voltage proportional to the digital input. The output voltages may be expressed mathematically by the following

$$V_1 = \left(\frac{x}{2^n - 1}\right) V$$

$$V_2 = \left(1 - \frac{x}{2^n - 1}\right) V$$

where x = the decimal equivalent of the binary input represented by the energizing of the relays.

The accuracy of this circuit depends upon the precision of the resistors in the voltage divider network and the relative impedance of the connected load. If the load impedance is not high compared with the network resistance, a cathode-follower stage may be introduced between the network and the load.

Design Problems

Two-Phase Wound Stators

The input voltage magnitudes to these windings may be approximated as shown in Figure 7 or 8.

The accuracy of the method of Figure 7 is ±4.1 degrees while that of Figure 8 is ±0.4 degree. The trapezoidal approximation of Figure 7 is therefore suitable for 5-digit or 6-digit coding where the degree of coding precision is ±5.6 degrees or ±2.8 degrees. Although the method of Figure 8 leads to a more complicated switching arrangement, its inherent accuracy makes it suitable for 8-digit or 9-digit coding where the degree of precision is ±0.7 degree or ±0.35 degree. It is apparent that for both of these methods some special switching must take place to select a given octant and that linear voltage dividers may be used inside the octant. The circuit diagram of a switching arrangement actually used for the multislope approximation of Figure 8 is given in Figure 9. This diagram shows the two stator coils

Table I

Binary Number	Corresponding Shaft Position in Degrees	Binary Number	Corresponding Shaft Position in Degrees
0 0 0 0 0	0	1 0 0 0 0	180
0 0 0 0 1	11¹/₄	1 0 0 0 1	191¹/₄
0 0 0 1 0	22¹/₂	1 0 0 1 0	202¹/₂
0 0 0 1 1	33³/₄	1 0 0 1 1	213³/₄
0 0 1 0 0	45	1 0 1 0 0	225
0 0 1 0 1	56¹/₄	1 0 1 0 1	236¹/₄
0 0 1 1 0	67¹/₂	1 0 1 1 0	247¹/₂
0 0 1 1 1	78³/₄	1 0 1 1 1	258³/₄
0 1 0 0 0	90	1 1 0 0 0	270
0 1 0 0 1	101¹/₄	1 1 0 0 1	281¹/₄
0 1 0 1 0	112¹/₂	1 1 0 1 0	292¹/₂
0 1 0 1 1	123³/₄	1 1 0 1 1	303³/₄
0 1 1 0 0	135	1 1 1 0 0	315
0 1 1 0 1	146¹/₄	1 1 1 0 1	326¹/₄
0 1 1 1 0	157¹/₂	1 1 1 1 0	337¹/₂
0 1 1 1 1	168³/₄	1 1 1 1 1	348³/₄

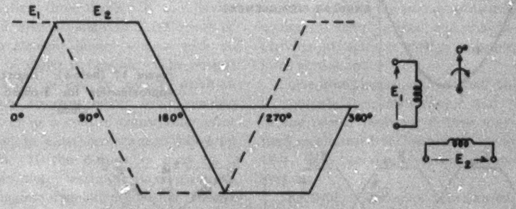


Figure 7. Sine wave approximation by trapezoids

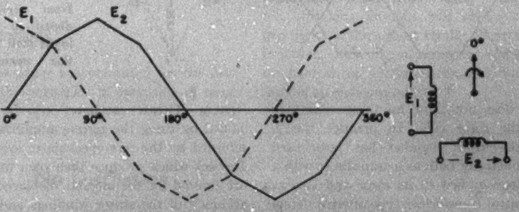
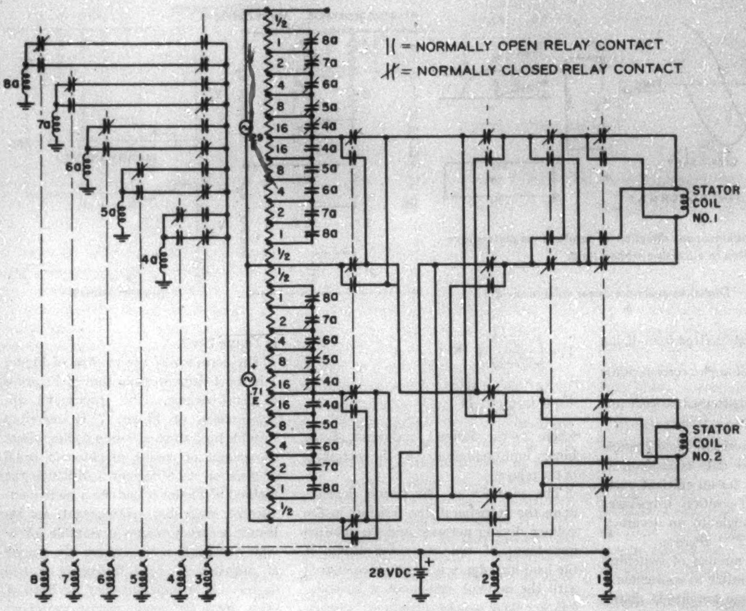


Figure 8. Sine wave approximation by multislope wave



)(= NORMALLY OPEN RELAY CONTACT
-)/(- = NORMALLY CLOSED RELAY CONTACT
½
1
2
4
8
16
16
8
4
2
1
½
8a
7a
6a
5a
4a
29
71
E
STATOR COIL NO. 1
STATOR COIL NO. 2
28VDC
+
8
7
6
5
4
3
2
1

Figure 9. Digital-to-analogue shaft position transducer

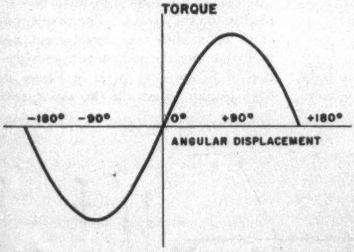
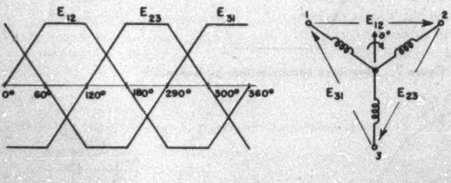


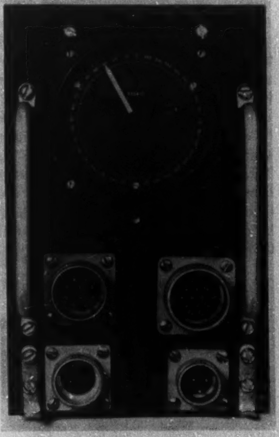
Figure 10 (left). Output torque versus angular displacement

Figure 11 (below). Trapezoidal approximation for 3-phase synchros



E_{12}
E_{23}
E_{31}
0°
60°
120°
180°
290°
300°
360°
1
2
3
E_{12}
0°
E_{31}
E_{23}

Figure 12 (right). Front view of digital-to-analogue shaft position transducer



but does not show anything connected to the rotor coil. In practice the 2-phase output member has been used satisfactorily both as a transducer with a voltage applied to its rotor and also as a control transformer type of error detector for a servomechanism follow-up system. Greater accuracy can be obtained by using the torque amplification afforded by the servomechanism system. Devices which produce their own torque are limited by the angular displacement accuracy of the stator windings and the fact that the torque versus angular displacement curve is sinusoidal as in a synchro (see Figure 10) and coulomb friction produces some errors. In practice the application of pulsating direct voltages to the stator windings of transducers reduced the effects of coulomb friction and also resulted in a lower operat-

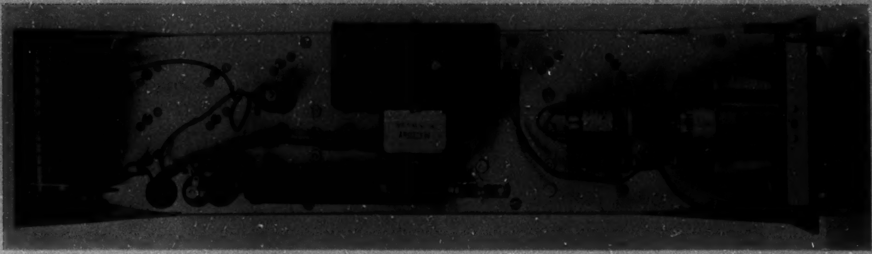


Figure 13. Top view of digital-to-analogue shaft position transducer

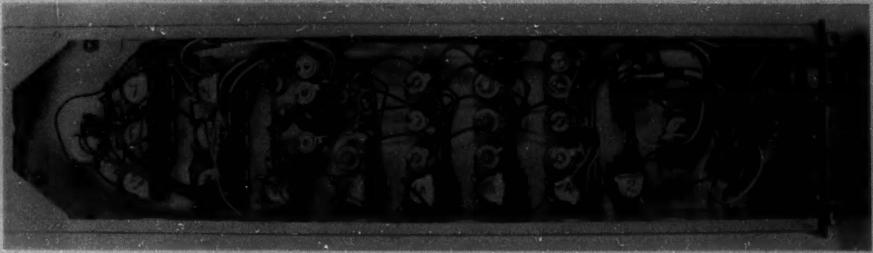


Figure 14. Bottom view of digital-to-analogue shaft position transducer

ing temperature.

The linear voltage approximations produce some unbalance between induced rotor voltages and the applied rotor voltages. These voltage differences set up circulating currents whose energy must be dissipated in the internal impedance of the voltage sources, the impedance of the transducer windings, and the voltage divider impedances. This internal dissipation of energy causes some heating, but this heating was not found to be excessive. The actual unbalanced rotor voltages are a function of the angular displacement of the transducer. A judicious choice of applied rotor voltage will limit this unbalance to equal maximum swings in either direction. For the trapezoidal case of Figure 7 this amounts to ±17 per cent of the applied voltage. For the multislope case of Figure 8 the maximum unbalanced voltage is ±3 per cent of the applied rotor voltage.

THREE-PHASE WOUND STATORS

The input voltage magnitudes to these windings may be approximated as shown in Figure 11.

The accuracy of this method is ±1.1 degrees which makes it suitable for 7-digit angle coding. A special form of coding is required since angular sextants, rather than octants, are used in this method. The first three digits may be used to select a sextant thereby wasting two out of eight possible combinations of the first three digits. If this device is used over an angle of 120 degrees or less, a straight binary system with no waste of digits is possible. The output of this device can be used to drive an ordinary synchro either as a synchro motor or as a control transformer error detector in a servomechanism follow-up system. If the output is used as a synchro motor to supply its own torque, the maximum unbalanced rotor voltage will be ±12 per cent of the applied rotor voltage.

Conclusions

This paper has presented for the first time the principles of operation of some minor-arc-sensing magnetic devices for converting from a digital parallel register output to a shaft position. The basic principles behind the generation of suitable alternating or direct voltages for application to standard 2-phase resolvers or 3-phase synchros have been developed.

Design approximations which can be used to minimize the amount of switching equipment were outlined, and criteria for arriving at suitable design approximations were shown.

These transducers can be used either to generate their own torque and turn to the proper position or they may be used as control transformers in conjunction with servomechanism follow-up systems.

The field of endeavor represented by the development of these digital-to-analogue transducers and their associated switching circuits will bridge the gap between the almost totally electronic functions of the digital computer and the mechanical or electromechanical functions of an analogue output.

No Discussion

Principles of Tape-to-Card Conversion in the AMA System

WILLARD B. GROTH
ASSOCIATE MEMBER AIEE

BILLING data for customer-dialed local and toll telephone messages are now recorded and processed by means of the automatic message accounting (AMA) system in the areas of seven principal cities in the Bell System network.

The AMA system records the information required for billing the customer in the form of perforations in paper tape in the central office. Recordings are made while the switching machinery is establishing connections between telephone lines and while the messages are in progress.

The perforated paper tapes are cut daily and forwarded to an accounting center where the information is processed through a series of machines each of which, with the exception of the last, produces new paper tapes. Individual message records are assembled and chargeable message units or chargeable time in the case of toll messages are determined. Individual message entries are sorted by calling line number, message units for a given billing period are added, and finally message unit summaries and individual customer toll slips are printed. The printed output of the AMA accounting center is used by clerical personnel in producing toll service statements and customer bills.

The AMA system in its first phase was completely described in an earlier paper.[1] It is the object of this paper to describe the latest development in the AMA system, namely the means and technique for converting at the most suitable stage in the process from the use of AMA paper tape to the use of standard business machine cards as a medium for handling message data. This development was carried on with the objective of making the accounting process more efficient by preparing toll service statements automatically and by providing the means for ultimately processing message unit data automatically. This process is known as tape-to-card conversion and will have a marked effect upon the routines followed in the accounting center. Tape-to-card conversion therefore is not merely an adjunct of the AMA system but a new process intended to supplement AMA in its present form. The system may then be more widely applied in an ever-increasing field.

Need for Improved Methods

Automatic message accounting, in its first phase, is not fully automatic in that the printed end products, message unit summary lists and toll slips, require manual processing. A message unit summary list for each central office indicates the summary of local message units for which each customer is to be billed. A portion of such a list with an interpretation of the numbers is shown in Figure 1. A toll slip shows for an individual customer a detailed record of all toll messages (nine per slip maximum) for a given billing period. A toll slip with the record for two toll messages is shown with an interpretation of the data in Figure 2.

A clerical force determines, from the message unit summary list, the charge including tax for each customer. They also compute the charge for each of the toll messages, the total of such charges, and the tax for all toll message business. A toll service statement giving details of all toll messages is furnished and a bill is prepared.

That message accounting on a fully automatic basis will eventually become imperative for efficient operation is indicated by the volume of traffic involved and the fact that an increasing proportion of toll messages, including the long-haul variety, will be customer-dialed in the not too distant future.

As an illustration of the tremendous volume of business involved, certain message rate statistics for the Bell System are shown in Table I. It might be expected that the numbers found in this table will increase through the coming years, and further, that the percentage of messages which are candidates for customer dialing will increase. This increased percentage might be anticipated since, with a spread in the use of customer dialing and the decrease in times for establishing connections using the new switching systems, the tendency will be toward more frequent use of station-to-station service.

Modifying the System

The plan for making the system more fully automatic with the prospect of eventually making it completely automatic is to use standard punched card machines, products of the International Business Machines (IBM) Corporation. A system plan was devised to arrange a variety of card machines to process toll message information for the production of toll service statements. The toll service statement lists detailed information for each toll message including charge, total toll message tax, and total charge plus tax for all toll messages for a billing period. This card machine system and its output are described in a companion paper.[2] Another system plan is to be made to use card machines for processing message unit message information.

The link required for bridging the gap between AMA and card machine systems is a machine which transfers message information from AMA tapes to business machine cards. The machine developed for converting from one recording medium to the other, for both local and toll messages, is the tape-to-card converter.

A brief review of the machine operations performed in AMA first phase will be given at this point so that the reader may gain a better understanding of the part played by the tape-to-card converter in the over-all system. The flow chart of

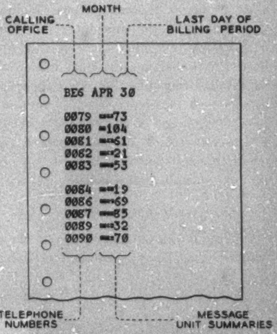


Figure 1. Portion of AMA message unit summary list and interpretation

---

Paper 53-113, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted November 19, 1952; made available for printing December 8, 1952.

WILLARD B. GROTH is with the Bell Telephone Laboratories, Inc., New York, N. Y.

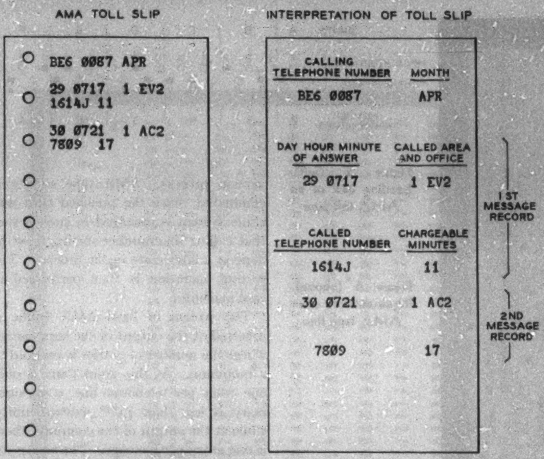

AMA TOLL SLIP
BE6 0087 APR
29 0717 1 EV2
1614J 11
30 0721 1 AC2
7809 17
INTERPRETATION OF TOLL SLIP
CALLING TELEPHONE NUMBER
MONTH
BE6 0087
APR
DAY HOUR MINUTE OF ANSWER
CALLED AREA AND OFFICE
29 0717
1 EV2
CALLED TELEPHONE NUMBER
CHARGEABLE MINUTES
1614J
11
1 ST MESSAGE RECORD
30 0721
1 AC2
7809
17
2ND MESSAGE RECORD

Figure 2. AMA toll slip and interpretation

Figure 3 indicates the first phase operations in sequence.

A central office AMA tape from a single recorder comes to the accounting center with a record of all messages, since the last tape cutting, that have passed over a group of a maximum of 100 trunk circuits. The record for each of the messages appears as several transverse lines of perforations in three distinct tape entries. Each of the entries is perforated at a definite stage in the progress of the call. The initial entry is made immediately after customer dialing and trunk seizure for the recording of calling line number and a 2-digit number designating the trunk. The called line number, called office, and area are also recorded if the message is to be detail-billed. The second entry is made when the called line is answered and records the start time of conversation and the number of the trunk. The third entry records the disconnect time and again the trunk number. Since many messages may be in progress at one time, the three piece records thereof may be interspersed on the common record tape.

The AMA assembler in the accounting center is used in a 2-stage process, sorting first by units and then tens digits of trunk number to perform the function of assembling the pieces of the individual message records on a new piece of AMA paper tape. Actually ten new tapes, numbered 0 to 9, are produced at each digital sort and they must be manually spliced together in numerical sequence to form a new continuous tape in preparation for the next stage in processing.

The AMA computer is then used to

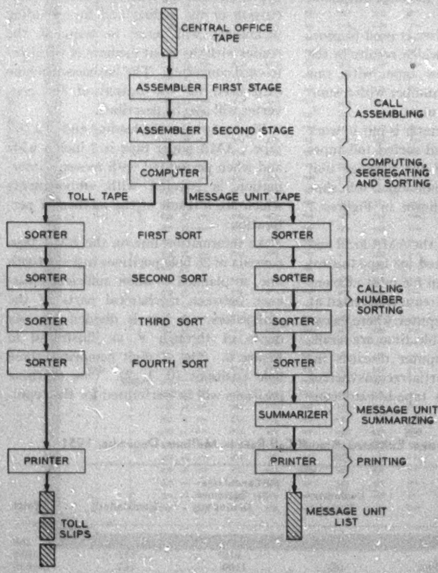

CENTRAL OFFICE TAPE
ASSEMBLER
FIRST STAGE
ASSEMBLER
SECOND STAGE
CALL ASSEMBLING
COMPUTER
COMPUTING, SEGREGATING AND SORTING
TOLL TAPE
MESSAGE UNIT TAPE
SORTER
FIRST SORT
SORTER
SORTER
SECOND SORT
SORTER
SORTER
THIRD SORT
SORTER
SORTER
FOURTH SORT
SORTER
CALLING NUMBER SORTING
SUMMARIZER
MESSAGE UNIT SUMMARIZING
PRINTER
PRINTER
PRINTING
TOLL SLIPS
MESSAGE UNIT LIST

Figure 3. Flow chart of AMA accounting center processing (first phase)

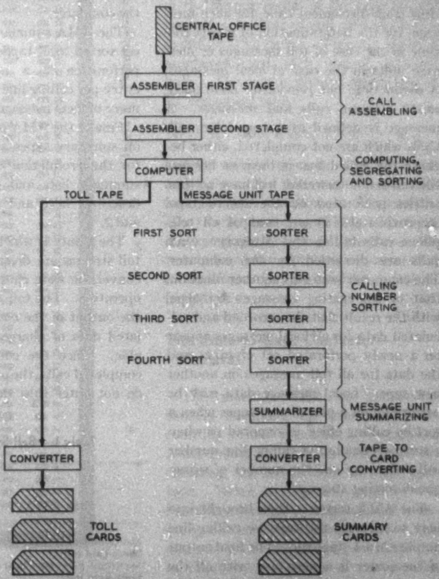
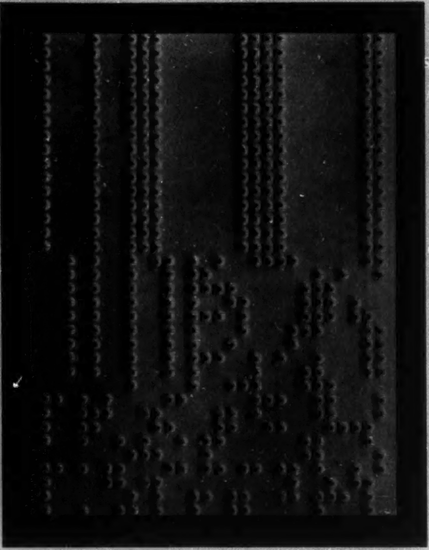

CENTRAL OFFICE TAPE
ASSEMBLER
FIRST STAGE
ASSEMBLER
SECOND STAGE
CALL ASSEMBLING
COMPUTER
COMPUTING, SEGREGATING AND SORTING
TOLL TAPE
MESSAGE UNIT TAPE
FIRST SORT
SORTER
SECOND SORT
SORTER
THIRD SORT
SORTER
FOURTH SORT
SORTER
CALLING NUMBER SORTING
SUMMARIZER
MESSAGE UNIT SUMMARIZING
CONVERTER
CONVERTER
TAPE TO CARD CONVERTING
TOLL CARDS
SUMMARY CARDS

Figure 4. Flow chart of AMA accounting center processing (with conversion)



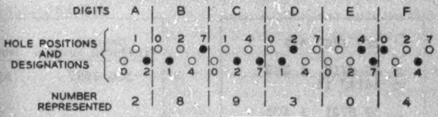


Figure 5 (left). Leading end of an AMA toll tape

Figure 6 (above). Representation of an AMA tape line

perform the function of subtracting start time from disconnect time, for each message, for the determination of chargeable time in the case of toll messages or message units in the case of local messages. A distinction, not previously mentioned, exists between calls and messages. A message is defined as a completed call. Calls which are not completed, either because the called line is busy or because the call is not answered, will have no time entries perforated on the record tape. Approximately 30 per cent of all telephone calls fall in this category. Such calls are discarded by the computer. The computer performs another function, that of segregating messages by type, with the result that the recorded and calculated data for all local messages appear on a newly perforated set of tapes and the data for all toll messages on another new tape. Local message data may be perforated on a plurality of tapes when a sort by calling office is required or when a sort by one digit of calling line number will reduce by one the number of subsequent sorting stages.

The AMA sorter is then brought into play to sort the toll tape by calling line number in a 4-stage run. The final output of the sorter is a new tape with all the messages for one telephone line in sequence and these calling line records in numerical sequence. The sorter run is repeated for the local message output of the computer.

The AMA summarizer is used to process sorted local tape which results in the perforation of a new tape with one entry per calling line number with a summary of local message units.

Finally the AMA printer is put to work on summary tapes and sorted toll tapes for the production of the message unit summary lists and individual toll slips described here and shown in Figures 1 and 2.

The points at which the AMA local and toll streams are diverted for tape-to-card conversion were chosen for most efficient operation. The toll stream is diverted at the output of the computer where calculated data of chargeable time are available. Since the computer discards incompleted calls, the partial records thereof do not enter into the tape-to-card conversion process. AMA tape sorting is eliminated, since the punched card machine system is arranged in such a way that calling line number sorting must be done at a later state in the process. The sorting operation is then performed in card machines.

The stream of local AMA traffic is diverted at the output of the summarizer where the number of entries is reduced to a minimum. At this point there is only one entry per telephone line, if the summary is less than 1,000 message units, while at the output of the computer there is one entry per message. The sequences of operations for local and toll business thus determined are portrayed in the flow chart of Figure 4.

AMA Tapes and Business Machine Cards

A brief description of the physical appearance of an AMA tape and a discussion of the coding and tape reading techniques will serve to acquaint the reader with the input medium of the tape-to-card converter. The business machine card used for the output of the converter will also be described.

Figure 5 shows the leading end of a toll tape. AMA paper tape is 3 inches wide and when perforated with message information is covered with embossments which are a result of the method of perforation.

An information line on the AMA tape consists of 28 hole positions in a sawtooth line, so planned to allow sufficient clearance between mechanical parts of the perforators. A line is divided into six digits, *A* through *F*, as illustrated in Figure 6. The *A* digit comprises three hole positions (0, 1, 2). One of these positions will be perforated for the repre-

Table I. Bell System Estimated Annual Call Rate in Millions, December 1951

Type of Call	Customer-Dialed	Candidates for Customer Dialing	Remainder	Total
Multiunit				568
Short-haul toll		970*	112	1082
Multiunit plus short-haul toll†	405	1100	145	1650
Long-haul toll		300*	300	600
Multiunit plus all toll		1805*	445	2250

* Includes those already customer-dialed.

† 5 cents to 35 cents.

sentation of a number. For example, the black circle in the *A* digit in Figure 6 indicates a perforation representing the number 2. Each of the remaining digits, *B* through *F*, comprises five hole positions (0, 1, 2, 4, 7). Two of these hole positions will be perforated in each digit for the representation of a number. Since there are ten combinations of five things taken two at a time, any one of the ten digital values, 0 to 9, may be represented in any of the digit positions, *B* through *F*. The hole positions are designed to represent numbers on an additive basis. For example, 1 and 7 in the *B* digit represent the number 8; 2 and 7 in the *C* digit represent the number 9 and so on. The only exception to the additive rule is the combination, 4 and 7, which is used to represent 0.

Adjacent lines on an AMA tape are dovetailed together so that the spacing between hole positions in a line and between lines is uniform.

An AMA reader has 28 steel sensing pins which are brought against the paper tape for the reading of a line. Pins encountering paper will be restricted in their motion while those finding holes will travel farther causing contact closures. Eleven contact closures constitute the reading of a good line.

A business machine card, representative of the output media of the tape-to-card converter, is made of stiff card stock $3^{1}/_{4}$ inches by $7^{3}/_{8}$ inches. Figure 7 shows a numbered card with several punched holes. The card is divided into a grid of 12 horizontal rows

12, 11, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9

and 80 vertical columns, each of which may be punched with one or more holes for the representation of information. The two uppermost positions, 12 and 11, are reserved for special punches; the remainder, 0 through 9, shown numbered in the illustration, are used for digital information.

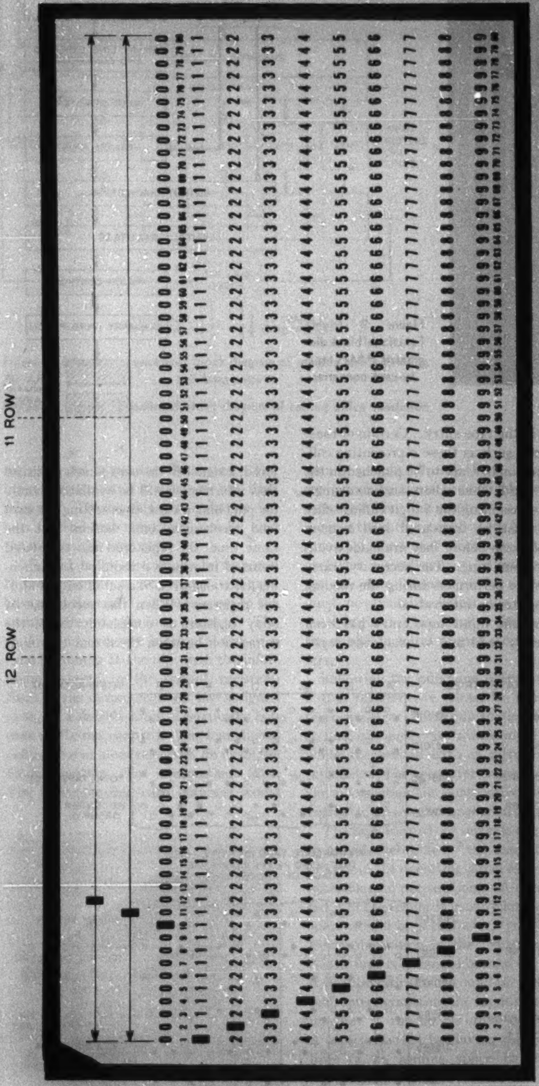


Figure 7. A numbered business machine card

The Tape-to-Card Converter—General

The machine used for conversion broadly resembles the other AMA accounting center machines. A standard AMA reader is used for electrically deciphering the information on the input tape. Relay circuits are used for the functions of memory, checking, and translating but here the similarity ends. The quantity and arrangement of the input information, the nature of the translations required, and the form of the output, as well as the manner in which the output mechanism operates, require the use of

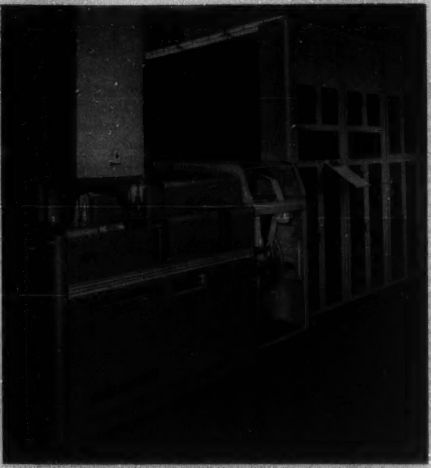


Figure 8 (left). The AMA tape-to-card converter

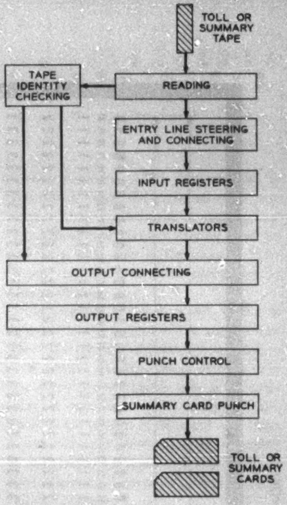

TOLL OR SUMMARY TAPE
TAPE IDENTITY CHECKING
READING
ENTRY LINE STEERING AND CONNECTING
INPUT REGISTERS
TRANSLATORS
OUTPUT CONNECTING
OUTPUT REGISTERS
PUNCH CONTROL
SUMMARY CARD PUNCH
TOLL OR SUMMARY CARDS

Figure 9 (right). Functional block diagram of AMA tape-to-card converter

unique circuitry throughout the machine.

The other AMA machines, with the exception of the printer, use perforators to produce output tapes of the same type as the input. The perforators are controlled by and synchronized with the reader, while the pattern of holes for a particular line is determined generally by information in relay storage registers. The converter, however, uses for its output a standard summary card punch. The card punch is controlled by relay circuits in the converter portion of the machine. The operation of these relay circuits in turn is controlled by and subject to the timing of the AMA reader. The control of the overall machine therefore is vested in the AMA reader, while the timing, to take full advantage of the speed of the output machine, is controlled by the fixed card cycle time of the punch. This problem of control and timing will be discussed in greater detail later.

Figure 8 shows the converter. Three relay cabinets, housing a total of approximately 700 relays and featuring operating and display panels, are shown with the reader cabinet and the associated IBM *519* Summary Punch.

Figure 9 is a functional block diagram of the converter. A tape to be processed for card conversion is fed into the reader of the tape-to-card converter. The job of the converter is to read and memorize the consecutive lines of a tape entry, perform certain translations, and control the card punch machine to punch memorized and translated information on a card.

The numbers are memorized exactly as found in the tape entry. Certain of these numbers, such as those representing calling line number, are to be punched in the output card, while others, such as a single digit representing the 3-digit calling office code, must be translated into a more suitable form before they are made available for punching. The necessary translations are performed during the reading and registering interval.

When the input tape entry has been completely read and satisfactorily registered, a signal to the card punch indicates that information will be available shortly for punching a card, thus setting the card feed mechanism into motion. At the same time, the registered and translated items of information required for punching are transferred to a set of output storage registers. When this second set of relay registers is completely filled, the connection between them and the input

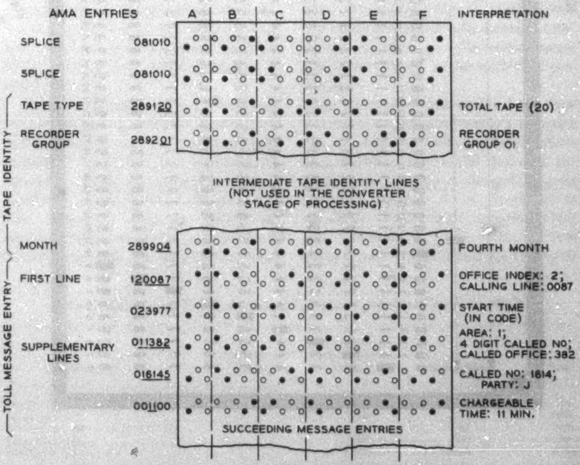
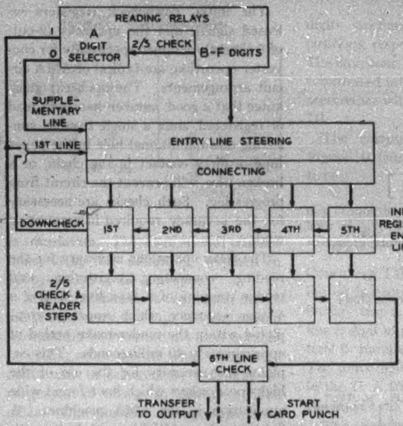

AMA ENTRIES
A
B
C
D
E
F
INTERPRETATION
SPLICE
081010
SPLICE
081010
TAPE IDENTITY
TAPE TYPE
289120
TOTAL TAPE (20)
RECORDER GROUP
289201
RECORDER GROUP 01
INTERMEDIATE TAPE IDENTITY LINES (NOT USED IN THE CONVERTER STAGE OF PROCESSING)
MONTH
289904
FOURTH MONTH
TOLL MESSAGE ENTRY
FIRST LINE
120087
OFFICE INDEX: 2; CALLING LINE: 0087
023977
START TIME (IN CODE)
SUPPLEMENTARY LINES
011382
AREA: 1; 4 DIGIT CALLED NO; CALLED OFFICE: 382
016145
CALLED NO: 1814; PARTY: J
001100
CHARGEABLE TIME: 11 MIN.
SUCCEEDING MESSAGE ENTRIES

Figure 10. Representation of leading end of typical AMA toll tape



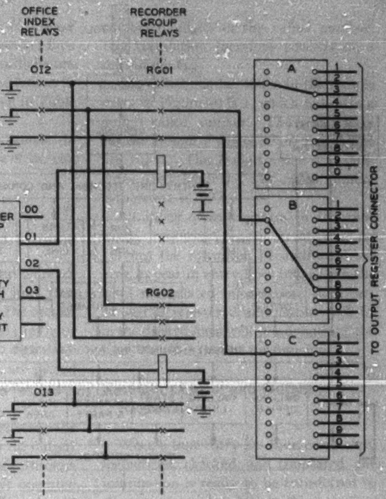


Figure 11 (above). Functional block diagram of reading and registering a toll tape entry

Figure 13 (right). Functional block diagram of calling office translation

registers and translators is severed. The input registers and translators are then released. The machine immediately begins to read and memorize, one by one, the lines in the next entry.

Meanwhile the card punch, in response to the first input register signal, will have advanced the card feed mechanism to the point where a card is to be punched. The card is punched, one horizontal row at a time beginning at the top, with the information stored in the output registers. Each of the output registers has assigned to it, on a flexible basis, a particular column on the output card. The four-digital calling line number registers, for example, might contain the information 1234. The corresponding card columns would then be punched during the 1, 2, 3, and 4 row punching times respectively to indicate the calling line number.

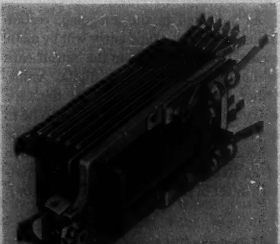


Figure 12. The general-purpose relay used in the AMA system

Since the card is punched one horizontal row at a time and since the punching cycle for a card cannot be interrupted, the complete information for a message must be available before the punching cycle begins. This dictates the need for the complete set of output storage registers.

Following the punching operation, the output registers are released. The registered and translated information for the second call entry is transferred thereto while the second card is advanced for punching. The converter thus deals with the information for consecutive message entries in an overlap operation.

The tape-to-card converter is controlled in actual operation by the manipulation of keys in the operating panel and certain other manual preparations. A tape to be fed into the reader is usually long enough to be wound on a reel which is placed in the feed position in the reader cabinet. As the tape is advanced beyond the reader it is automatically rewound on a take up reel, thus permitting the input side of the machine to run on an unattended basis. A tape is started by having the first few lines of its leading splice pattern advanced, under manual control, to engage the reader drum. A machine start key causes the reader to read and advance consecutive lines of splice pattern until the first of a group of lines known as tape identity is encountered. Figure 10 shows a representation of the leading end of an AMA toll tape and an interpretation of the numbers thereon. The machine checks tape identity items against the settings of manually controlled rotary switches on the operating panel. If there is agreement, the machine steps the tape through tape identity. Following tape identity, the machine encounters message entries and proceeds to convert the information, entry by entry, in overlap operation as described previously, to run the card punch at the rate of 100 cards per minute.

Trailing tape identity follows the last message entry and the machine rechecks to insure that switches have not been reset inadvertently during the run. Resetting of a switch might result in false accounting. The reader then skips trailing splice pattern and an alarm is sounded to indicate the end of the run.

Minor modifications in the IBM *519* Summary Punch together with a special circuit arrangement in the converter permit stopping and restarting the machine from the card punch operating panel. This facilitates the manual operation of card removal. Blank card supply may be replenished while the machine is running.

A more detailed description of the operation involved in several broad functions of the converter is given now. This will serve to illustrate the manner in which switching circuit techniques have been applied to the field of automatic message

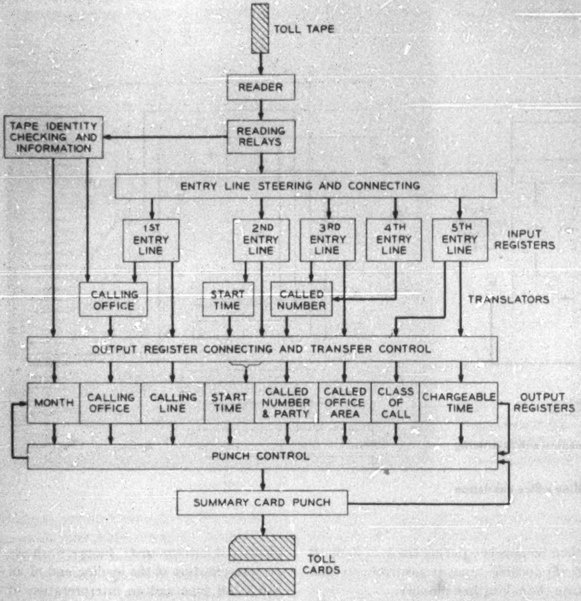

TOLL TAPE
READER
TAPE IDENTITY CHECKING AND INFORMATION
READING RELAYS
ENTRY LINE STEERING AND CONNECTING
1ST ENTRY LINE
2ND ENTRY LINE
3RD ENTRY LINE
4TH ENTRY LINE
5TH ENTRY LINE
INPUT REGISTERS
CALLING OFFICE
START TIME
CALLED NUMBER
TRANSLATORS
OUTPUT REGISTER CONNECTING AND TRANSFER CONTROL
MONTH
CALLING OFFICE
CALLING LINE
START TIME
CALLED NUMBER & PARTY
CALLED OFFICE AREA
CLASS OF CALL
CHARGEABLE TIME
OUTPUT REGISTERS
PUNCH CONTROL
SUMMARY CARD PUNCH
TOLL CARDS

Figure 14. Functional block diagram of AMA tape-to-card converter, detailed for toll use

accounting, and at the same time give a better understanding of the tape-to-card converter.

Reading and Registering a Tape Entry

A toll message entry with an interpretation of the numbers therein is represented in Figure 10. This sample entry may be used as a reference in the ensuing descriptions.

Figure 11 is a functional block diagram of the reading and registering operations. The first step in reading and registering a tape entry is to establish a connection between the reading relays and the first line input register. The connection is made subject to the coincidence of two electrical checks:

1. All input line registers released (this signal is known as a downcheck).
2. Reading relays, *B* through *F* digits, operated on a 2-out-of-5 basis plus a 1 in the *A* digit (first line indication).

When the connection is made, the first line register is operated to store the first tape entry line. Following this, a contact network on the first line register relays will pass a signal, if the relays are operated on a 2-out-of-5 basis (such a signal is known as a 2-out-of-5 check or an upcheck), to permit the reader to step, advancing the tape by one line. The reader step signal also advances a steering circuit in preparation for making the next connection to the second line register.

A new 2-out-of-5 check on the *B* through *F* digit reading relays plus a 0 in the *A* digit (supplementary line indication) causes this second connection to be made. The first and supplementary line indications may be found in the *A* digits of the toll entry in Figure 10.

This steering, connecting, and stepping sequence is used to read and register the five consecutive lines of the tape entry in the five input line registers. It should be noted that only the *A* digit of a line need be known to the machine for determining its action in these operations. The information in the *B* through *F* digits of each of the five lines is memorized by the device of locking relays operated.

The initial downcheck (registers released signal) and the upchecks (2-out-of-5 checks), designed specifically for converter operation, are typical of AMA circuit arrangements. The upchecks guarantee that a good number has been read or registered, since a single trouble, consisting of an additional hole in the paper tape, a dirty contact in the chain, or a broken wire, will prevent the circuit from progressing. Such checks are necessary for the accuracy required in accounting work.

The relay operations necessary for the reading, connecting, registering, and reader stepping of a tape line involve a 4-stage sequence which must be completed within the reader make period of approximately 40 milliseconds. This explains the necessity for the use of the high-speed relays which have found wide application in the AMA machines. A flat type relay, of fairly recent design with upper and lower sets of card operated twin bifurcated contacts, has been used as the work horse in the AMA system. One of these relays, which has an operate time of approximately 7 milliseconds, is illustrated in Figure 12. During the break or open portion of the reader cycle, which lasts approximately 20 milliseconds, the reading, connecting, and steering relays release simultaneously.

When the first line of the second entry is read, 1 in the *A* digit indicates that the correct number of supplementary lines has been registered for the first entry. This indication, plus the five input line register upchecks, initiates two signals:

1. Signal to start card punch for punching first message card.
2. Signal to transfer input registration plus translations to output registers for punching.

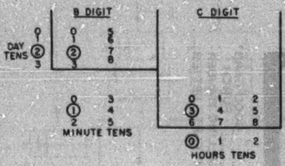
Translation in the Tape-to-Card Converter

As indicated in the foregoing, certain items in the input tape entry will be made available for punching on the output card without any change in digital value. Other items of information in the tape entry must undergo some change to be in a presentable form for punching. The function of changing the representation of an item of information from one pattern or set of numbers to another is known

Table II. Numbers and Letters in Telephone Dial Holes

1	2	3	4	5	6	7	8	9	0
	ABC	DEF	GHI	JKL	MNO	PRS	TUV	WXY	Z

Table III. Start Time Code



as translation. There are several translations, using this definition, required in the operation of the converter.

In one translation, the converter makes use of tape entry and tape identity switch information to construct a 3-digit number for calling office identification. Any three digits may be selected, by a flexible arrangement of wiring, to identify a given office. One possible arrangement is to translate to the digits that would be used in dialing that office. This translation is described now.

The *B* digit of the first line of a toll entry is the office index, which indicates the number of the calling office within a group of central offices served by a common group of AMA recorders, a common situation in metropolitan areas.

The office index, and the recorder group number as indicated by the tape identity switches, identify the calling office. The block diagram of Figure 13 shows in abbreviated form the manner in which the translation to calling office code is accomplished. The combination of an office index relay operated from the first line *B* digit, 2 in this example, and a recorder group relay corresponding to the settings of the recorder group switches, 01 in this case, passes a signal to each of three terminal strips, *A*, *B*, and *C*. Cross connections to the opposite sides of these strips perform the translation which is extended to the output registers when the connectors are operated. The calling office code for the connections shown might be *EV*2. The information, however, is placed in the output registers on a numerical basis, which for this example would be 382. The corresponding numbers and letters in the dial holes are shown in Table II.

Another translation must be made using certain numbers in the second line of the tape entry. This line contains numbers giving the start time which consists of day tens and units, hour tens and units, and minute tens and units. In the interest of confining the start time to one line, the computer is designed to compress the six digits into the five digit spaces of a tape line. This is accomplished by placing the day, hour, and minute units digits, unchanged, in the *D*, *E*, and *F* positions, respectively, of the second line. The day, hour, and minute tens digits are compressed into the *B* and *C* positions in accordance with the code shown in Table III.

The computer, for the example in Figure 10, found the start time of the call to be:

day of month	29
hour of day	07
minute of hour	17

Consulting Table III, it is found by reading on horizontal and vertical lines that for day tens 2, minute tens 1, the second line *B* digit must be 2. Similarly for hour tens 0, minute tens 1, the *C* digit is 3. The units digits in order, 977, are placed in the *D*, *E*, and *F* positions, thus giving the start time line, 23977.

The converter must re-expand the start time line for punching on the output card. The translation is made through networks of contacts on relays operated directly from the *B* and *C* digit second line input register relays. The translation to day, hour, and minute tens is rendered on a decimal basis, in preference to 2-out-of-5 code, in order to limit the number of contacts used.

A third rearrangement of numbers, which may be considered to be a special kind of translation, is concerned with called number. The called number is found in the fourth line of the entry and may consist of four or five digits. In addition a 4-digit number may be followed by a party letter.

The highest digit of called number, whether it be thousands or ten thousands, is found in the *B* position of the fourth line with lower place digits to the right in positions *C*, *D*, and so forth. When a party letter is associated with a 4-digit number it is placed in the *F* position. The party letters *J*, *M*, *R*, and *W* are represented by the corresponding dial hole numbers, 5, 6, 7, and 9 respectively. When a party letter is not used with a 4-digit number the *F* position is blanked with zero.

The converter determines the number of digits to be used by means of the *C* digit of the third line. This digit, designated called number index, may assume values of 1 or 2 to indicate a 4- or 5-digit called number. The called number index determines which of two connectors shall be operated at the time that information is transferred to the output registers. The connectors are designed to place ten thousands, thousands, hundreds, tens, units and party numbers in correspondingly designated output registers, irrespective of the number of digits in the called number. The translation therefore consists of a shift of digits.

The called number index in the sample entry of Figure 10 is 1, which indicates a 4-digit called number. The numbers in the *B* through *F* digits of the fourth line are 16145. The *F* digit value of 5 indicates party letter *J*. The called number is therefore 1614*J*.

All items of information in the tape entry, requiring special treatment involving the operation of additional relays, appear in entry lines other than the last. Translated information, therefore, as well as registered information, will be in the desired form well in advance of the time that it is needed for punching.

Transfer Control and Output Registration

When a tape entry has been completely memorized, checked, and translated, the information is ready to be transferred to the output registers for punching.

Figure 14 is a functional block diagram of the converter, which includes the input registers and translators and indicates, by arrows and the names of the various output registers, the items of information contributed by each of the input sources. It is the job of the transfer control circuit to determine when the transfer of information from input to output shall take place. The transfer control establishes connections between input and output upon the coincidence of two electric signals:

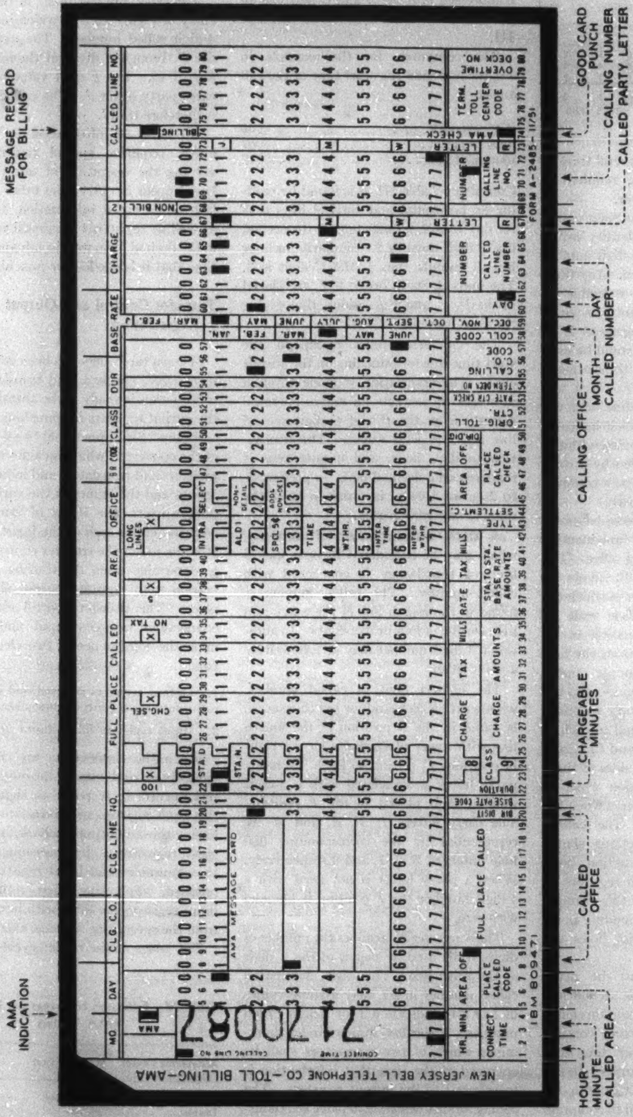
1. Output registers released and ready for call information (output downcheck).
2. Input registers filled (input upcheck).

When the connections are made between input and output, the output registers operate in response to signals from the input registers and translators. Following this an output upcheck, insuring a good registration for punching, causes the connectors and input registers to be released. Following their release, the input registers are filled with information from the succeeding message entry.

The times of the reader cycle and the

Table IV. Estimated Messages per Month in AMA Centers

Accounting Center	Message Unit	Toll
Philadelphia	3,500,000	130,000
Detroit	3,900,000	50,500
Newark	4,035,000	3,350,000
Chicago	650,000	690,000
New York	1,070,000	
San Francisco	305,000	



MESSAGE RECORD FOR BILLING
AMA INDICATION
GOOD CARD PUNCH
CALLING NUMBER
CALLED PARTY LETTER
DAY
CALLED NUMBER
MONTH
CALLING OFFICE
CHARGEABLE MINUTES
CALLED OFFICE
HOUR
MINUTE
CALLED AREA
NEW JERSEY BELL TELEPHONE CO.—TOLL BILLING-AMA
AMA MESSAGE CARD
IBM 809471
FORM A-2405-11/51
7170087

Figure 15. A punched toll card

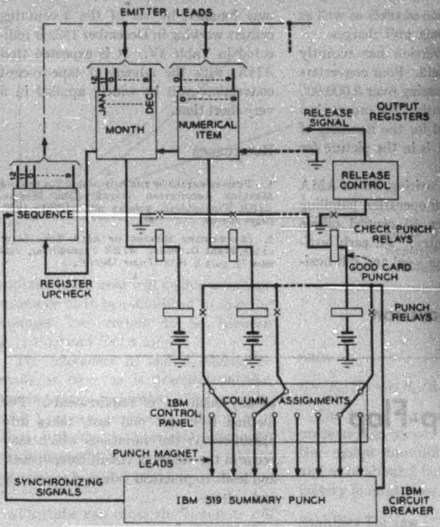


Figure 16 (left). **Functional block diagram of punch control**

Figure 17 (above). **IBM control panel with typical toll wiring**

punch machine card cycle are such that the input will be refilled before the completion of the punching operation. Since the transfer operation is made very quickly, new information will be available for almost immediate use after each punching cycle. The punch machine will therefore run continuously.

Punching the Output Card

Figure 15 shows a typical punched card for toll message processing. The items of information are those of the sample entry of Figure 10. The information in Figure 15 may also be compared with the details of the first entry on the toll slip of Figure 2; they are equivalent.

The punching of the card is made possible by the sequential grounding of 12 emitter leads, one for each horizontal row of punching, after an output upcheck has been detected. Figure 16 is a functional block diagram of the punch control, which shows the relationship of the emitter leads, the output registers, and the card punch. When emitter 12 lead is grounded, a check punch and punch relay will be operated in series through each register storing a 12 indication. This operation takes place shortly before the punching time and thus provides a preset pattern for the punching of the 12 row on the card. Punching is accomplished by the operation of the corresponding punch magnets, which remove interposers under the punching dies, upon the closure of the IBM circuit breaker. At the end of 12 punching time, emitter lead 11 is grounded and the punch relays which were operated for the 12 row are released. The card, moved in synchronism with the emitter pulses, is then advanced to place the 11 row under the punching dies. The remainder of the card is punched row by row in this fashion.

The construction of the check punch relays is such that they remain operated until knocked down by a signal at the end of the card-punching cycle. If a check punch relay is operated for every register, a path will be closed at 9 punching time to operate the good card punch. Thus an electrical and a visual check are provided to insure that the card has been correctly punched. The good card punch signal releases the output registers after 9 time, completing the punching cycle.

The IBM control panel, indicated in Figure 16, provides a means of plug wiring for column assignments on a flexible basis. Figure 17 shows an IBM control panel with typical toll wiring.

After a card is advanced in the summary punch beyond the punching stage, it is sensed to insure that the correct number of punches has been made. This is called double punch and blank column detection and may be assigned to columns on a flexible basis through the IBM control panel.

The card may be end-printed with quarter-inch characters in the following station of the summary card punch. Connect time and calling line number, selected by control panel wiring, are shown end printed at the left side of Figure 15. These items are used for visually selecting certain message records. The end printing completes the over-all tape-to-card conversion process.

Summary of Converter Circuit Functions

The detailed description of the tape-to-card converter has been given in connection with the handling of toll business. The broad circuit functions are similar for the processing of local message unit summaries, although the number of lines to be registered per entry and the translations to be performed are fewer.

It is interesting to note that each of the broad circuit functions of the converter, when closely examined, makes use of one or more fundamental switching circuit concepts. The novelty of the circuits, of course, is in the contact configurations, the sequences of operations, and the particular applications. The function of reading and registering, for example, makes use of steering, connecting, memory, and checking circuits. Following registration, translating circuits are employed. Checking, connecting, and memory circuits are used again in the transfer to output registers. A sequence circuit is used for generating the emitter pulses. Finally the arrangement for operating punch relays through output registers, some of which are operated on a 2-out-of-5 basis and others on a decimal basis, constitutes a translation to special and digital punch card codes.

Conclusion

The benefits of tape-to-card conversion are manifold. The process, supplemented by a system of standard punched card machines, eliminates the costly and time-consuming manual operations of ticket sorting, toll rating, toll billing, toll adding, and determination of taxes as well as the handling of message unit charges.

Tape-to-card conversion has recently been applied in the field. Four converters are collectively processing over 3,000,000 toll messages per month. The process has not as yet been used for the handling of local business, but it is in the picture for the very near future.

There are, at this writing, seven AMA accounting centers in operation handling the business of over 50 central offices throughout the country. An early estimate of the amount of local and toll business handled by six of the accounting centers working in December 1952 is indicated in Table IV. It is expected that AMA, with the advent of tape-to-card conversion, will be widely applied in a very short time.

References

1. Fundamentals of the Automatic Telephone Message Accounting System, **John Meszar.** *AIEE Transactions*, volume 69, part I, 1950, pages 255–69.

2. Mechanized Billing of AMA Toll Messages, **Frank D. Slade.** *AIEE Transactions*, volume 72, part I, 1953 (*Paper 53-114*).

No Discussion

A Graphical Method for Flip-Flop Design

R. F. JOHNSTON
NONMEMBER AIEE

A. G. RATZ
MEMBER AIEE

Synopsis: A method of design is described for flip-flops subject to any set of requirements and examples of the method are given. The method is simple and fast. It takes into consideration those variations which may occur in the values of circuit components and leads to practical values of resistance and voltage.

THE Eccles-Jordan trigger circuit, or flip-flop, is employed extensively in digital computers, binary counters, scalers, and similar devices. Some applications use a large number of flip-flops under conditions where they must function satisfactorily for long periods of time; special design techniques are required to ensure the high reliability required.

The flip-flop circuit[1] shown in Figure 1 is essentially a symmetrical 2-stage overdriven d-c amplifier coupled back on itself in such a way that it locks into one of two stable states. Suppose T_1 conducts and T_2 does not. The points *a* and *d* are therefore at lower potentials than the corresponding points *b* and *c*. If the resistors are chosen properly, the potential of *d* is sufficiently negative to keep T_2 cut off, while the potential of *c* is above ground and T_1 continues to conduct. Thus the circuit settles into a state in which T_1 conducting renders T_2 nonconducting and T_2 nonconducting permits T_1 to conduct. It also can settle into a state with T_2 conducting and T_1 cut off.

Stability, that is, the continued ability of the circuit to settle in either of its two stable states when suitably triggered, is the most important characteristic of a flip-flop. In addition, the flip-flop may have to meet certain other requirements; for example, the magnitude of the voltage difference between states at a plate or grid may have to have a specified value, or the transition from one state to the other may have to be sufficiently fast. It is also essential that performance does not depend critically on any circuit quantities.

The method of design described in this paper can be applied to a flip-flop subject to any such set of requirements. The method is simple and fast, takes into consideration the variations which may occur in the values of circuit components, and leads to practical values of resistance and voltage.

Design Tolerances and Tube Characteristic Safety Factors

The quantities involved in the design are the resistances R_1, R_2, R_3, the voltage levels E_1, E_2, and the characteristics of the tube. (T_1 and T_2 are usually the two halves of a double triode.) Because variations in these quantities affect the circuit in different ways, two types of safety factors must be used.

Tolerances

A resistor can cause the circuit to fail if its resistance is either too large or too small. Considering manufacturer's tolerance and drifts due to temperature, aging, and so forth, it is usually possible to find a quantity Δ_r such that all resistors of a given make are certain to have a resistance lying between $R(1-\Delta_r)$ and $R(1+\Delta_r)$, if R is the nominal resistance. The design must be arranged to meet specifications with resistors lying anywhere in this range.

The power supplies can cause the circuit to fail if either E_1 or E_2 is too large or too small. The characteristics of the

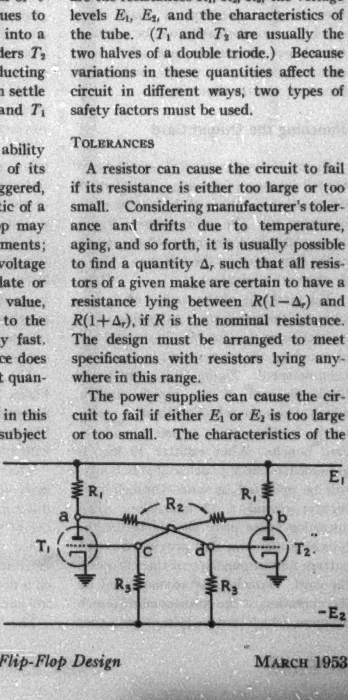


Figure 1. Basic flip-flop circuit

Paper 53-60, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted April 14, 1952; made available for printing December 3, 1952.

R. F. Johnston and A. G. Ratz are at the University of Toronto, Toronto, Ontario, Canada.

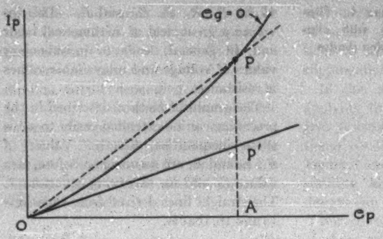


Figure 2 (left). Tube characteristics

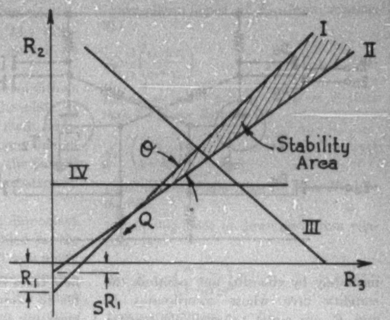


Figure 3 (right). Representation of inequalities

supplies to be used will enable a voltage tolerance Δ_e to be established so that the voltages are certain to lie between $E_1(1 \pm \Delta_e)$ and $E_2(1 \pm \Delta_e)$.

The tolerances Δ_r and Δ_e should be made as large as is practical without making design impossible. For example, with 1-per-cent resistors, a value of Δ_r of about 0.05 (that is, 5-per-cent tolerances) might be used.

TUBE FACTORS

The tube can cause the circuit to fail in two ways. First, if the current drawn by the conducting tube is too small, the other tube will not be cut off. A good design ensures that the grid of the conducting tube is positive with respect to the cathode (grid-cathode current limits the difference to 1 or 2 volts), so that the zero-grid plate characteristic determines conservatively the current drawn by this tube. Figure 2 shows this characteristic for a typical triode. The curve is almost linear and can be approximated by the straight line OP. As the tube ages, the slope of OP decreases; the limit of active life in the flip-flop is determined by the decrease which can be tolerated. The flip-flop should be designed using a zero-grid characteristic such as line OP'; the larger the ratio: $r = (AP/AP')$, the longer the tube life. The conducting tube can then be represented as a resistor

$$R_t = (OA/AP') = r\left(\frac{OA}{AP}\right) \tag{1}$$

Secondly, the circuit can fail if too great a negative grid voltage is required to keep the nonconducting tube cut off. This cutoff voltage is proportional to plate voltage, which usually is not known at the beginning of a design; it is certainly safe to assume that the full value of E_1 is applied at the plate. Thus, if $(-E_{co})$ is the grid cutoff obtained from the characteristic curves for any particular plate voltage equal to E_1, the ratio (E_{co}/E_1) is assumed constant for that tube type. Since E_{co} increases with tube life, the design must be based on a ratio

$$e_c = (E_c/E_1) = m(E_{co}/E_1), \quad m > 1 \tag{2}$$

The safety factors r and m should be made as large as possible without making the design too difficult. Values of r and m of about 2 have been found satisfactory in most cases.

Stability Design

Every flip-flop must satisfy two stability conditions for all values of resistances and voltage levels within the tolerance ranges previously defined.

CONDITION 1

To keep the grid of the conducting tube positive, it is necessary that

$$\frac{(R_1+R_2)}{R_3}\,\frac{(1+\Delta_r)}{(1-\Delta_r)} < \frac{E_1}{E_2}\,\frac{(1-\Delta_e)}{(1+\Delta_e)} \tag{3}$$

or

$$R_2 < g\left(\frac{E_1}{E_2}\right)R_3 - R_1 \tag{4}$$

where

$$g = \left(\frac{1-\Delta_e}{1+\Delta_e}\right)\left(\frac{1-\Delta_r}{1+\Delta_r}\right) \tag{5}$$

$$\simeq 1 - 2(\Delta_e + \Delta_r) \tag{6}$$

CONDITION 2

To keep the nonconducting tube cut off it is necessary that

$$\frac{\left(\dfrac{R_tE_1}{R_t+R_1(1-\Delta_r)} + E_c\right)(1+\Delta_e)}{E_2(1-\Delta_e) - E_c(1+\Delta_e)} < \frac{\left(\dfrac{R_tR_1}{R_t+R_1(1-\Delta_r)} + R_3\right)(1-\Delta_r)}{R_3(1+\Delta_r)} \tag{7}$$

or

$$R_2 > \frac{1}{g}\left(\frac{s+e_c}{\dfrac{E_2}{E_1} - e_c}\right)R_3 - sR_1 \tag{8}$$

where

$$s = \frac{R_t}{R_t + R_1} \tag{9}$$

(While equation 8 derives from equation 7 by an approximation, the difference is negligible if the values of r and m exceed about 1.1.)

A design is obtained by solving inequalities 4 and 8. To do this, tube, resistor, and power supply characteristics must be known well enough to determine Δ_r, Δ_e, R_t, and e_c. If values are then assumed for R_1 and (E_1/E_2), relations 4 and 8 reduce to linear inequalities in R_2 and R_3.

The usual method of solving these relations[2] is to replace the inequality signs by equality signs and to solve the resulting pair of equations. This is equivalent to plotting

line I

$$R_2 = g\left(\frac{E_1}{E_2}\right)R_3 - R_1 \tag{10}$$

line II

$$R_2 = \frac{1}{g}\left(\frac{s+e_c}{\dfrac{E_2}{E_1} - e_c}\right)R_3 - sR_1 \tag{11}$$

and reading the co-ordinates R_2, R_3 of the point of intersection (point Q in Figure 3). In general, the values of R_2 and R_3 obtained in this way are not practical and it is necessary to test the nearest available values by substitution in relations 4 and 8 to ensure that these relations are satisfied.

Properly interpreted, however, Figure 3 is a graphical representation of inequalities 4 and 8: all points below line I satisfy equation 4, all points above line II satisfy equation 8 and all points in the shaded region, called the stability area, satisfy both relations 4 and 8. Thus a practical design is obtained im-

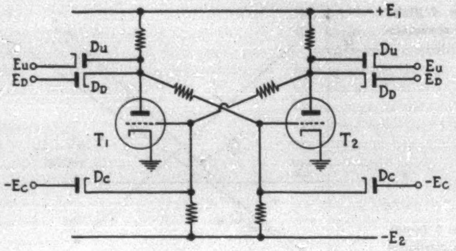


Figure 4. Flip-flop with clipping diodes

mediately by choosing any point in the stability area whose co-ordinates R_2 and R_3 are equal to available resistor values.

If the initial assumptions regarding (E_1/E_2) and R_1 do not lead to a satisfactory design, the effects of new assumptions are determined easily through their effects on lines *I* and *II* in Figure 3. Anything that is done to move the point *Q* downward in the direction of the arrow or to increase the angle between the lines makes it easier to find satisfactory values of R_2 and R_3.

The switching speed of a flip-flop depends to a considerable extent on the value of R_1; for a high-speed circuit R_1 should be small.[3] By equating the slopes of equations 10 and 11, the smallest possible value of R_1 is seen to be

$$\mathbf{R}_1 \doteq R_t\left[\frac{1+\Delta_r}{g^2\left(1-e_c\frac{E_1}{E_2}\right)-e_c}-1\right] \tag{12}$$

However, if R_2 and R_3 are to be reasonably small, it is generally required that

$$R_1 \geq 2\mathbf{R}_1 \tag{13}$$

Equation 12 shows that R_1 is particularly sensitive to R_t, Δ_r, and Δ_e and somewhat to e_c. If a tube is available which is capable of passing more current or which has a lower value for e_c, or if resistors or power sources of better stability can be used, a considerable reduction can be made in R_1.

In the same way, it can be shown that the minimum possible value for E_1/E_2 is

$$\frac{E_1}{E_2} > \frac{1}{e_c}\left(1-\frac{s+e_c}{g^2}\right) \tag{14}$$

A value of E_1/E_2 in the range 0.5 to 2 is generally satisfactory.

General Design

Besides the basic stability conditions discussed so far, many flip-flop applications require that voltages and impedances throughout the circuit satisfy additional conditions. While only the more common cases will be treated here, it is possible to extend this graphical method to meet any set of specifications.

The most common examples of such additional conditions concern plate voltage swings. It is often required that the difference between the voltage at the plate of the nonconducting tube (the up-plate) and that at the plate of the conducting tube (the down-plate) shall not be smaller than a given minimum value ΔE_M. This is so if:

(1) The up-plate is always more positive than a voltage E_U; that is, if

$$R_2 > \left(\frac{E_U}{E_1}\right)\left(\frac{R_1}{g-\frac{E_U}{E_1}}\right) \tag{15}$$

(2) The down-plate is always more negative than a voltage E_D; that is, if

$$R_2+R_3 < \frac{g(sR_1)\left(\frac{E_2}{E_1}+\frac{E_D}{E_1}\right)}{s-g\frac{E_D}{E_1}} \tag{16}$$

(3) $|\Delta E_M| \leq E_U - E_D$ (17)

Extending the first method mentioned under the section entitled "Stability Design," a design could be obtained by solving the five (nonlinear) equations derived from equations 4, 8, 15, 16, and 17 for R_1, R_2, R_3, E_1, and E_2. This involves a great deal of arithmetical labor and, in general, leads to inconvenient values of voltage and unavailable values of resistance.

The graphical method described in the preceding can be extended easily to solve these inequalities directly. Values of E_1, E_2 and R_1 are assumed as before, also values E_U and E_D satisfying equation 17. The straight lines derived from equations 15 and 16, that is,

line *III*

$$R_2 = \left(\frac{E_U}{E_1}\right)\left(\frac{R_1}{g-\frac{E_U}{E_1}}\right) \tag{18}$$

line *IV*

$$R_2+R_3 = \frac{g(sR_1)\left(\frac{E_2}{E_1}+\frac{E_D}{E_1}\right)}{s-g\frac{E_D}{E_1}} \tag{19}$$

are plotted on the same graph, see Figure 3, as the stability area obtained from equations 4 and 8. Then the co-ordinates R_2, R_3 of any point in the stability area, above line *III* and below line *IV*, satisfy all the conditions. The choice of a point whose co-ordinates equal available resistor values completes the design.

In some circuits the up-plate must be exactly at a voltage E_U; this can be ensured by designing as previously for the up-plate to be more positive than E_U, then adding a clipping diode from this point to a fixed voltage E_U. Similarly, the down-plate may be clipped at E_D or the down-grid at $-E_c$; one diode must be used at each point for each voltage level at which it is desired to clip.

Figure 4 shows a flip-flop complete with clipping diodes at both plates and grids. Diodes D_D are used to limit the excursions of the down-plates to E_D, diodes D_U hold the up-plates to E_U and diodes D_c

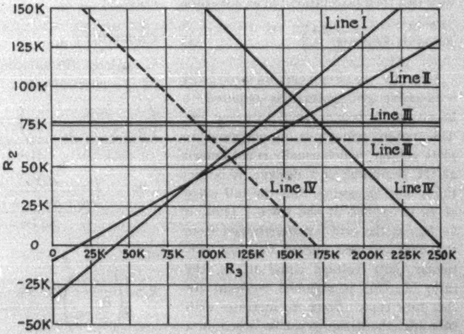


Figure 5. Illustrating the examples in Appendix II

hold the down-grids to $-E_c$. (Note that the up-grid is always clipped by grid current if relation 4 is satisfied.)

There are many variations of the basic flip-flop circuit depending on which of D_D, D_U, and D_c are used. Some of these require changes in stability conditions 4 and 8. For example, to satisfy the stability condition at the up-grid of a flip-flop which uses diodes D_U, equation 4 must be altered to

$$R_2 < g\frac{E_U}{E_1}\frac{E_1}{E_2}R_3 \tag{20}$$

Equation 20 is then used instead of equation 4 in finding the stability area in Figure 3. If diodes D_D are used, equation 8 becomes

$$R_2 > \frac{1}{g}\left(\frac{\dfrac{E_D}{E_1}+e_c}{\dfrac{E_2}{E_1}-e_c}\right) \tag{21}$$

It is[1] noteworthy that the use of diodes D_U increases the tube cutoff safety factor m. In the definition of m, equation 2, E_1 was used as a conservative estimate of the plate voltage of the nonconducting tube. If the up-plate is clipped, however, the plate voltage cannot exceed E_U, and the effective value of m is increased by the ratio (E_1/E_U).

In still other applications of flip-flops, it is desired to use a voltage tapped off the transfer resistor R_2. If this voltage is E_T, then it is required that

$$E_T \geq E_{TU} \tag{22}$$

for one state of the flip-flop and that

$$E_T \leq E_{TD} \tag{23}$$

for the other state.

If R_2 is tapped at a point such that there is a resistance fR_2 $(f<1)$ between the grid end of R_2 and the point, leaving a resistance $(1-f)$ R_2 between this point and the plate, then relation 22 is satisfied if

$$R_2 > R_1\left[\frac{\dfrac{E_{TU}}{E_1}}{fg-\dfrac{E_{TU}}{E_1}}\right] \tag{24}$$

and relation 23 is satisfied if

$$R_2 > \left|\frac{sR_1 - xR_3}{fx-(1-f)}\right| \tag{25}$$

where

$$x = \frac{1}{g}\left[\frac{s-\dfrac{E_{TD}}{E_1}}{\dfrac{E_2}{E_1}+\dfrac{E_{TD}}{E_1}}\right] \tag{26}$$

All points in the stability area above the horizontal line derived from relation 24 and above the line derived from relation 25 satisfy relations 22 and 23. Relations 24 and 25 are correct only if no clipping diodes are used.

In this way, too, voltage conditions applying to any points on the plate or grid resistors can be met. Each additional condition adds one more line in Figure 3 and further restricts the part of stability area from which satisfactory designs can be obtained.

For convenience, the most important of these relations are assembled in the summary of Appendix I. To simplify the expressions and calculations, a change in notation has been made: all voltages are divided by E_1 and the resulting "normalized" voltage ratio represented by a small letter, thus

$$\frac{E_U}{E_1} = e_U$$

Appendix II shows two examples of flip-flop design to illustrate the use of this method.

Appendix I. Summary of Design Relationships

Let

$$\frac{E_U}{E_1} = e_U, \quad \frac{E_D}{E_1} = e_D, \quad \frac{E_2}{E_1} = e_2, \quad \frac{E_c}{E_1} = e_c$$

Also

$$s = \frac{R_t}{R_t + R_1}; \quad g = \left(\frac{1-\Delta_r}{1+\Delta_r}\right)\left(\frac{1-\Delta_e}{1+\Delta_e}\right)$$

Stability Conditions at Up-Grid

No clipping diode at up-plate (from relation 4)

$$R_2 < \frac{g}{e_2}R_3 - R_1 \tag{27}$$

Up-plate clipped at E_U (from relation 8)

$$R_2 < g\left(\frac{e_U}{e_2}\right)R_3 \tag{28}$$

Stability Conditions at Down-Grid

No clipping diode at down-plate (from relation 20)

$$R_2 > \frac{1}{g}\left(\frac{s+e_c}{e_2-e_c}\right)R_3 - sR_1 \tag{29}$$

Down-plate clipped at E_D (from relation 21)

$$R_2 > \frac{1}{g}\left(\frac{e_D+e_c}{e_2-e_c}\right) \tag{30}$$

Condition that Up-Plate Be More Positive than E_U (from Relation 18)

$$R_2 > \left(\frac{e_U}{g-e_U}\right)R_1 \tag{31}$$

Condition that Down-Plate Be More Negative than E_D

No clipping diode at down-grid (from relation (19)

$$R_2 + R_3 < gs\left(\frac{e_2+e_D}{s-ge_D}\right)R_1 \tag{32}$$

Down-grid clipped at $-E_c$

$$R_3 < gs\left(\frac{e_c+e_D}{s-ge_D}\right)R_1 \tag{33}$$

Appendix II. Examples of Flip-Flop Design

To illustrate the application of the method described in Appendix I, an example is worked out.

Example I

Suppose it is desired to design a flip-flop using a *6J6* double triode for an application which requires high reliability and a time constant at the plate of less than 0.5 microsecond and assuming an output capacity of 10 micromicrofarads. Resistors in Radio Manufacturers Association sizes and having manufacturer's tolerance of ±1 per cent are to be used: the available power supplies have 200 volt outputs, accurate to ±1 volt.

PRELIMINARY DECISIONS

Tolerances: To ensure trouble-free performance during the life of the resistors take

$\Delta_r = 5 \times 1$ per cent $= 5$ per cent $= 0.05$

Since the power supplies are (nominally) accurate to 0.5 per cent, take

$\Delta_e = 10 \times 0.5$ per cent $= 5$ per cent $= 0.05$

Tube: For a *6J6:* from equation 2, $\frac{E_c}{E_1} = 0.10$ with $m=2$, and from equation 1, $R_t = 15K$ with $r=2$.

Choice of R_1: On the basis of the discussion

Table I. Table of Correct Combinations of Relationships

Required Conditions of Voltage at			Stability Conditions	Conditions Other Than Stability
Up-Plate	Down-Plate	Down-Grid		
Free	free	free $L-E_c$	27, 29	
Free $>E_U$	free	free $L-E_c$	27, 29	31
Free	free $<E_D$	free $L-E_c$	27, 29	32
Free $>E_U$	free $<E_D$	free $L-E_c$	27, 29	31, 32
Clipped at E_U	free $<E_D$	free $L-E_c$	28, 29	31, 32
Free $>E_U$	clipped at E_D	free $L-E_c$	27, 30	31, 32
Clipped at E_U	clipped at E_D	free $L-E_c$	29, 30	31, 32
Clipped at E_U	clipped at E_D	clipped at $-E_c$	28, 30	31, 33

concerning equations 12 and 13, choose

$R_1 = 33K$

(The time constant governing transition is then less than 0.3 microsecond.[3])

DESIGN

From the preceding values: $g = 0.82$; $s = 0.31$; $e_c = 0.10$; $e_2 = 1.0$. Substitution of these in relations 27 and 29 yields for line *I*

$R_2 < 0.82R_3 - 33$

and for line *II*:

$R_2 > 0.56R_3 - 10$

Note that all resistor values are expressed in thousands of ohms. These lines are plotted in Figure 5.

Any one of the points:

P_1: $R_2 = 47K$, $R_3 = 100K$
P_2: $R_2 = 62K$, $R_3 = 120K$
P_3: $R_2 = 75K$, $R_3 = 150K$
P_4: $R_2 = 87K$, $R_3 = 150K$

and many others would represent a satisfactory design. In general, faster transition will be obtained with smaller values of R_2 and R_3, so the choice of P_1 is indicated.

The design can be summarized thus: $E_1 = E_2 = 200$ volts; $R_1 = 33K$, $R_2 = 47K$, $R_3 = 100K$.

Example II

Suppose the flip-flop in the previous example also is required to have a plate swing of at least 50 volts.

DESIGN

The stability conditions for this example already are represented by the stability area in Figure 5. For this additional condition, relations 17, 31, and 32 must be used, with $E_M = 50$ volts.

A first assumption for E_U and E_D satisfying equation 18 might be

$E_U = 110$ volts; $E_D = 60$ volts

This yields for line *III*

$R_3 > 67K$

and for line *IV*

$R_2 + R_3 < 170K$

These lines are shown dashed in Figure 5; clearly there is no part of the stability area above this line *III* and below this line *IV*.

If the new assumptions are made

$E_U = 115$ volts; $E_D = 65$ volts

then line *III* becomes

$R_3 > 77.5K$

and line *IV* becomes

$R_2 + R_3 < 250K$

These lines are plotted on Figure 5 and it is seen that there is a part of the stability area, containing P_4, which satisfies all the conditions imposed. Thus the design derived from P_4 will satisfy the stability conditions and will guarantee a plate swing of at least 50 volts (from 65 volts to 115 volts) between states. The final design, then, is: $E_1 = E_2 = 200$ volts; $R_1 = 33K$, $R_2 = 87K$, $R_3 = 150K$.

References

1. ULTRA-HIGH FREQUENCY TECHNIQUES (book), **Brainerd, Koehler, Reich, Woodruff.** D. Van Nostrand, Inc., New York, N. Y.

2. PROGRESS REPORTS, **Signal Corps Electronic Computer Research and Development.** Moore School of Engineering, Philadelphia, Pa.

3. WAVEFORMS (book), **Chance, Vernon, MacNickol, Sayce, Williams.** McGraw-Hill Book Company, Inc., New York, N. Y.

No Discussion

Automatic Call Recording and Accounting in the SATT System

J. E. OSTLINE
FELLOW AIEE

Introduction

AUTOMATIC SWITCHING

When the first automatic telephone office was placed in public service at La Porte, Ind., in 1892, automatic switching and subscriber dialing became established facts. During the years following this notable event in the history of telephone communication, Automatic Electric Company through constant research and development of switching equipment and operating methods expanded the field of automatic switching from local exchange areas to suburban and toll networks. By 1910 operator toll dialing was introduced whereby originating toll operators were enabled to complete calls to automatic stations in distant towns by toll dialing.[1] Operator toll dialing was then expanded to provide switching of toll circuits through tandem and switching centers. During the period from 1920 to 1940, several regional automatic toll switching networks were established in various parts of the United States and Canada.

Although during that period operator toll dialing became an accepted method of toll switching, subscriber dialing of toll calls was not as yet available for several basic reasons, some of which were:

1. The numbering plans used by operators were not universal and therefore not suitable for public use.
2. If a universal numbering plan for toll switching had been available, common control switching equipment would have been required to make use of it.
3. All toll calls had to be recorded (ticketed) for billing purposes. This is one of the functions performed by toll operators.

THE DIRECTOR

With the introduction of the Director[2] in 1923 for the London, England, metropolitan area network universal numbering and common control operation became available for Strowger (step-by-step) automatic switching in local exchange networks. The Director also made available a number of other controls, and thus it became possible to design equipment suitable for subscriber dialing of short-haul and long-distance toll calls with automatic call recording.

SUBSCRIBER DIALED AUTOMATIC TOLL SWITCHING

In 1937, Régie des Télégraphes et des Téléphones (Belgian Telecommunication Authority) (R.T.T.) requested Automatique Electrique, S.A. (Belgian affiliate of the Automatic Electric Company) to present plans for subscriber national toll dialing with automatic recording (ticketing) of each dialed toll call. As a basic part of this undertaking, the R.T.T. had established a nation-wide area code numbering plan; see Figure 1.

The various common control facilities available in the Director provided the solution for all phases of this request, in-

---

Paper 53-111, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted November 19, 1952; made available for printing December 4, 1952.

J. E. OSTLINE is with the Automatic Electric Company, Chicago, Ill.

Grateful acknowledgment is made to my Automatic Electric Company associates and to R. F. Stehlik of Automatique Electrique S.A., Antwerp, Belgium, for suggestions and assistance in preparing this paper.

cluding automatic call recording (ticketing) with cost computation for each call; see Figure 2. The first installation of this system was made at Mons, Belgium, where it proved entirely satisfactory. A similar installation was placed in service at La Louviere in 1950.[3]

Further Developments

After the completion of the Belgian automatic toll ticketing plan in 1938, research was continued further to expand the automatic toll switching and call recording facilities contained therein. This resulted in a design which contained the following added facilities and methods of operation:

1. Detector (identifier) operation was arranged to provide subscriber dialing for party line subscribers up to five stations per line.
2. Central office toll call recording was arranged to permit the introduction of punched card accounting techniques, thereby making available complete automatic bill processing and accounting of toll messages.

Tee system resulting from the foregoing development is now referred to as the SATT (Strowger Automatic Toll Ticketing) system. When used with punched card accounting, this system is completely automatic from the dialing of toll calls by a subscriber to the printing of the bill for such calls.

A small trial unit containing most of the facilities in this system was installed in an office in the Los Angeles metropolitan network in 1944. Since that time

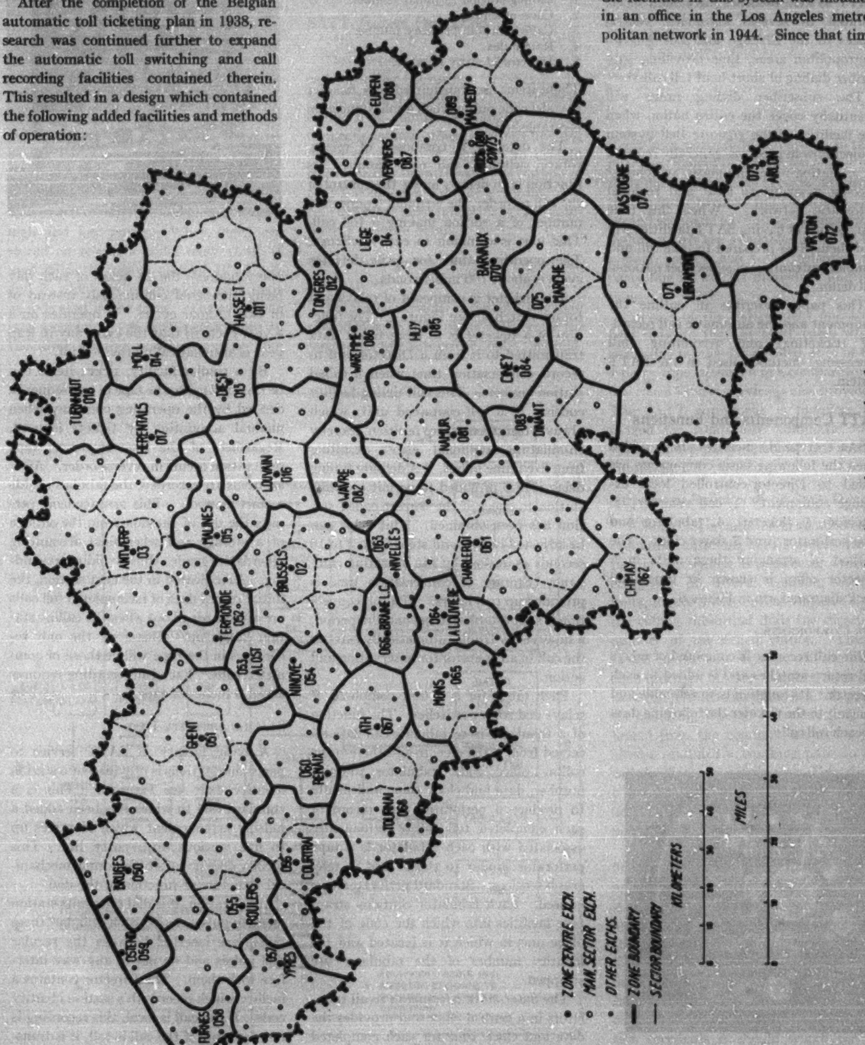
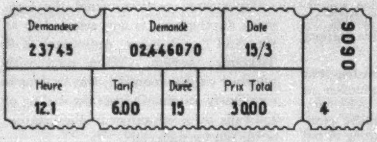

EUPEN 088
089 MALMEDY
080 PONTS
VERVIERS 087
ARLON 073
BASTOGNE 074
VIRTON 072
LIEGE 04
BARVAUX 070
071 LIBRAMONT
075 MARCHE
TONGRES 012
HASSELT 011
MOLL 014
DIEST 013
TURNHOUT 018
HERENTALS 017
WAREMME 086
HUY 085
CINEY 084
083 DINANT
NAMUR 081
LOUVAIN 016
WAVRE 082
MALINES 015
ANTWERP 03
BRUSSELS 02
063 NIVELLES
CHARLEROI 061
CHIMAY 062
TERMONDE 052
066 BRAINE L.C.
064 LA LOUVIERE
053 ALOST
NINOVE 054
MONS 065
GHENT 091
ATH 067
060 RENAIX
TOURNAI 068
BRUGES 050
056 COURTRAI
055 ROULERS
OSTEND 059
057 YPRES
FURNES 058
ZONE CENTRE EXCH.
MAN. SECTOR EXCH.
OTHER EXCHS.
ZONE BOUNDARY
SECTOR BOUNDARY
KILOMETERS
MILES

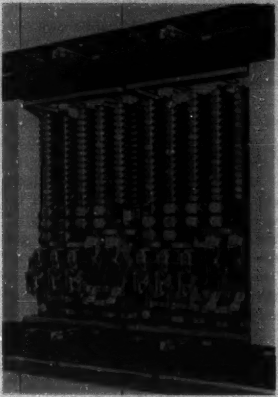
Figure 1. Map of Belgium showing national numbering plan areas with codes



Demandeur 23745	Demandé 02446070	Date 15/3	0909
Heure 12.1	Tarif 6.00	Durée 15	Prix Total 30.00
			4

Figure 2 (left). Automatic toll ticket

Figure 4 (right). Two ticketer units



SATT system facilities have been incorporated in several automatic central offices in the Chicago and Los Angeles metropolitan areas, thus providing subscriber dialing of short-haul toll calls.[4]

The subscriber dialing range will eventually cover the entire nation when the facilities of the gigantic Bell System project now in progress, generally referred to as nation-wide automatic switching,[5] become available to local and regional automatic networks. When this time arrives, offices having SATT facilities will find themselves prepared to take full and complete advantage of subscriber national toll dialing.

This paper describes in outline the equipment and the automatic call recording (ticketing) and accounting (bill processing) methods used in the SATT system.

SATT Components and Functions

SATT (type A) facilities are provided when the following basic components are added to Director-controlled local exchange equipment: 1. call recorder; 2. detector; 3. ticketer; 4. tabulator and tape perforator; and 5. dater-clock. The method of arranging these units in a Director office is shown in functional block diagram form in Figure 5.

THE COMPONENTS

The call recorder is composed of relays and rotary switches and is added to each Director. Its function is to assemble and transmit to the ticketer the following data for each toll call:

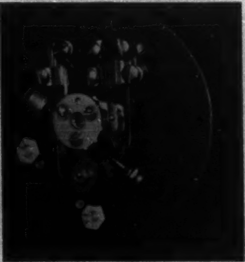


Figure 3. Identity dial

1. Calling station terminal number.
2. Class of service.
3. Called station directory number.
4. Rate index.
5. Director identity number.

These items are transmitted to a ticketer in codel code pulses when a toll connection is established.

The detector is composed of relays, rotary switches, and neon lamps. Its function is to detect, under the control of the call recorder, the terminal (connector) number of a station making a toll call. This unit is common to each exchange. The operating time for each detector cycle is about 100 milliseconds.

The ticketer is composed of relays and rotary switches; see Figure 4. The function of a ticketer is to store ticket data transmitted to it from a Director and to record conversation time when a called station answers. This call timing facility consists of a self-contained unit, which always starts from a zero position, thereby eliminating fractional errors resulting from fixed time pulses. A definite timing delay is also provided to permit a calling station to check if the correct called station has been obtained. This delay can be adjusted in 1-second steps from 3 to 10 seconds as desired by the operating company. Storage of conversation time is provided up to 5 hours. If a toll connection is kept established beyond this period, a supervisory trunk automatically extends the call to an operator for observation and action.

Each tabulator is a unit composed of relays and rotary switches. The function of a tabulator is to tabulate call data received from a ticketer, to add thereto the calling office code, tabulator identity number, date and clock time, and finally to produce a perforated tape record of each completed toll call. Permanently associated with each tabulator is a tape perforator similar to those used in telegraph services. Standard perforator tape is used. Each tabulator contains strapping facilities into which the code of the office unit in which it is located and the identity number of the tabulator are strapped.

The dater-clock is common to all tabulators in a central office and provides the date and clock time for each completed toll call. Clock time is expressed in Navy figures—the 24 hours of each day being numbered continuously instead of in two 12-hour cycles. It operates on a 4-year cycle so that the extra day in leap year is automatically provided for.

By a similar facility, ticket items can be transmitted to the tape in any sequence desired by the operating company when manual accounting by tickets is used. Regardless of the item sequence, tape perforation occurs in reverse order. After all items are recorded, the tabulator adds a start symbol. This arrangement permits the use of the tape from the outside of a storage reel when final accounting records are made, thus avoiding rewinding. As indicated in the next section, the initial record data of incompleted toll calls are always wiped out when the calling station hangs up. Therefore, the only recordings in the tape will be those of completed calls. Each call recording requires about 3 inches of tape.

STATION IDENTITY DIAL

A basic facility of SATT service to party-line stations is provided by a station identity dial; see Figure 3. This is a standard dial to which has been added a pulsing arrangement which identifies up to five stations on a party line. This facility does not alter the normal mechanical structure or function of the dial.

In operation, this dial transmits station identity pulses during each pull, but these pulses are inserted between the regular dial pulses and do not in any way interfere with them. The Director contains a facility which records this station identity code. If the call is local, this recording is not used, but if the call is toll, it is transferred to the call recorder, Figure 5.

Therefore, when a call recorder tests for a calling station terminal number, it already knows what service a calling station has, that is, individual line or party line. If party line, it knows which station of a maximum of five on the line has dialed the call. It will, therefore, make the terminal number detection (identification) test accordingly.

SATT System Code

All call data transmitted and recorded in the SATT system consist of numerals in a code pattern referred to as the codel code. Each digit is represented by combinations of four elements for transmission pulses, storage relays, or tape perforations as shown in Table I.

When data from the call recorder are pulsed out to the ticketer (see the next section), each digit is composed of four high and low pulses. When digits are stored on relays, a codel relay group is used with the relays marked *W*, *X*, *Y*, and *Z*, which are operated for each digit as indicated in the table. The tape code is a 4-channel type with perforations in the same relative positions as the high pulses in each code digit, Figure 8. When a tape is processed, the start symbol is used to start transfer of tape data into either a ticket printer (manual accounting) or into a card punch (automatic accounting). The blank symbol is used to fill in the third office code digit space when combined 2-4 and 2-5 numbering is used in an exchange area.

SATT System Operation

This section deals with the automatic call recording (ticketing) facilities in the SATT (type *A*) system. The methods of establishing subscriber dialed speech connections and other functions are not described except as required to explain call recording operations. The outline of SATT components in the preceding section describes most functions of each component in more detail.

Call Recording from Party-Line Stations

The initial versions of the SATT system provided subscriber dialed toll service for individual line subscribers only. As relatively large portions of subscribers in telephone exchanges in the United States use party-line service, it became apparent that in order to make full use of the service and economic advantages of subscriber dialing which are available in the SATT system, service had to be provided for party-line as well as individual line stations.

Table I. Functional Chart of Codel Code Pulsing, Storage, and Tape Recording

	Pulse Position				Relays Operated				Tape Channel			
Digit	1	2	3	4	W.	X.	Y.	Z.	1	2	3	4
1	H	H	L	L	*	*			.	.		
2	H	L	H	L	*		*		.		.	
3	H	L	L	H	*			*	.			.
4	L	H	H	L		*	*			.	.	
5	L	H	L	H		*		*		.		.
6	L	L	H	H			*	*			.	.
7	H	L	L	L	*				.			
8	L	H	L	L		*				.		
9	L	L	H	L			*				.	
0	L	L	L	H				*				.
Start									.	.	.	.
Blank										.	.	.

Surveys of the independent operating field indicate that in most cases where party-line service is provided, it is of the 4-party type, although in a comparatively few cases 5-party service is also provided. The SATT equipment is therefore arranged for completely automatic individual and 5-party terminal per station automatic call recording.

Automatic Recording of Call Data

When a call is initiated in an automatic exchange containing SATT system facilities, it is extended to a Director which is composed of several common control units, Figure 5. After receiving dial tone, the calling subscriber dials the directory number of the called station. This is stored in the call register. If the office code portion of this number indicates that the terminating office is a toll point, the resulting directive (translation) obtained from the common translator will always establish a trunking path to the called office via a ticketer. This directive will also alert the call recorder. This unit of the Director assembles the major portion of the toll call data.

The call recorder upon being alerted first of all receives from the call register an indication as to whether the calling station is on an individual or party line, and if on a party line, which particular station on that line has initiated the call. The call recorder now calls for the common detector (identifier). As soon as the detector responds, the call recorder informs the detector in which particular group of station terminals it should search for the calling station. In the case of an office

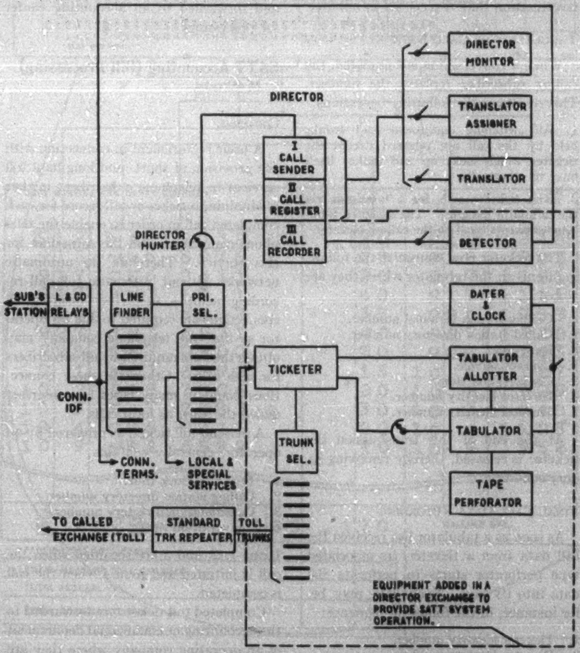


Figure 5. Central office SATT system functional block diagram

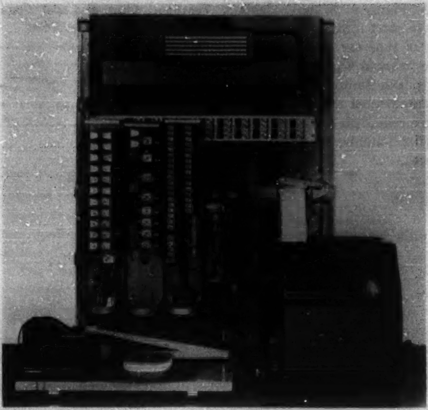


Figure 6 (left). Tape reader and ticket printer

Figure 7 (above). Card punch with tape-to-card control unit

providing individual and 5-party service, this means one of six different number groups. The detector then tests in the indicated group for the terminal number of the calling station, and transmits this number to the call recorder where it is stored. The detector then releases.

The call sender unit of the Director will be establishing the desired connection while the foregoing cycle in the call recorder takes place. As soon as the call recorder has dismissed the detector, it will start sending the ticket data now available in the Director to the ticketer where it is stored. This transmission, which is in codel code, takes place simultaneously with the establishing of the trunk (speech path) connection to the called office.

The call data transmitted to the ticketer from the call recorder consists of the following:

1. Calling station terminal number.
2. Called station directory number.
3. Rate index.
4. Class of service.

In addition to these call items, the call recorder also transmits the identity number of the Director handling the call.

After a connection has been established to the called central office terminal, the Director releases. The calling station is now connected to the called station connector terminal via a ticketer in the originating office, and the call (ticket) particulars accumulated up to this point are written up (recorded) in the ticketer. At this stage, the call may result in a busy signal or no answer. Should this occur, the calling subscriber replaces the handset, whereupon all equipment involved in the connection is released, and the accumulated call data are wiped out.

Timing of Conversation

When a call is answered by a called station, timing of conversation begins after a fixed delay to permit the calling subscriber to ascertain that the correct called station has been obtained. The conversation time is recorded in minutes.

Tabulation of Call Data

When a conversation is completed, the calling subscriber replaces the handset. This results in the following operations:

1. All switching equipment and trunks held by the call are released except the ticketer which locks up and makes itself busy to other calls.
2. The ticketer calls for a tabulator via the tabulator assigner. The assigned tabulator connects itself to the calling ticketer.

The ticketer now transmits the following items to the tabulator where they are stored:

1. Calling station terminal number.
2. Called station directory number.
3. Rate index.
4. Class of service.
5. Conversation time.
6. Director identity number.
7. Ticketer identity number.

At the end of this transmission the ticketer is released, thereby removing its busy condition.

Final Call Data Recording

As soon as a tabulator has received the call data from a ticketer, its associated tape perforator starts to perforate the data into its tape. Such data may be, for instance, in the following sequence:

1. Ticketer identity number.
2. Director identity number.
3. Date.
4. Clock time.
5. Called station directory number.
6. Conversation time.
7. Rate index.
8. Class of service.
9. Calling station directory number.
10. Tabulator identity number.

To this series of items is always added a start symbol. This is followed by the restoration to normal of all tabulator components thus making the tabulator available for another call. The punched tape is automatically wound up on a storage reel from which it is removed periodically and forwarded to an accounting center for processing.

SATT Accounting (Bill Processing) Methods

General

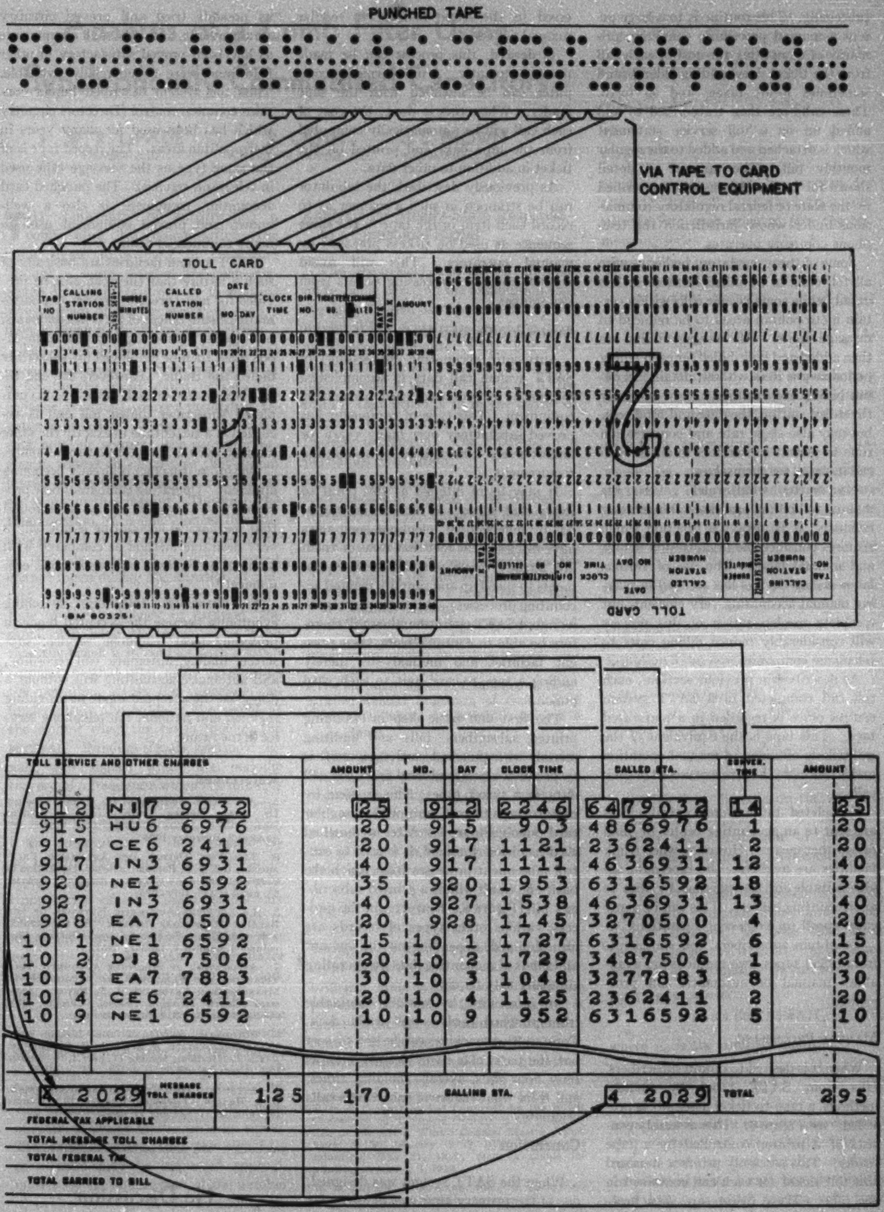
A basic requirement in connection with the provision of short- and long-haul toll services to telephone subscribers is that of obtaining a ticket or call record for each completed call in order to enable the telephone company to bill the subscriber for this service. Therefore, in automatic networks without automatic toll call recording facilities, subscribers desiring toll connections are required to dial an operator so that the telephone company may obtain the data required to bill subscribers for such calls. Otherwise these connections could in many cases be subscriber dialed the same as local calls.

A manual toll ticket, as prepared by an operator, generally contains:

1. Date and clock time.
2. Calling station directory number.
3. Called station directory number.
4. Conversation time.

Items 1, 2, and 3 are recorded when the call is initiated and item 4 when the call is completed.

Completed toll tickets are forwarded to the accounting or commercial department of an operating company where they are put through a routine referred to as bill


PUNCHED TAPE
VIA TAPE TO CARD CONTROL EQUIPMENT
TOLL CARD
TAB NO
CALLING STATION NUMBER
CLASS SERVICE
NUMBER MINUTES
CALLED STATION NUMBER
DATE
MO.
DAY
CLOCK TIME
DIR. NO.
TICKETER NO.
EXCHANGE CALLED
RATE
TAX %
AMOUNT
1
2
IBM 803251

TOLL SERVICE AND OTHER CHARGES			AMOUNT	MO.	DAY	CLOCK TIME	CALLED STA.	CONVER. TIME	AMOUNT
9 12	NI7	9032	25	9	12	2246	64 79032	14	25
9 15	HU6	6976	20	9	15	903	4866976	1	20
9 17	CE6	2411	20	9	17	1121	2362411	2	20
9 17	IN3	6931	40	9	17	1111	4636931	12	40
9 20	NE1	6592	35	9	20	1953	6316592	18	35
9 27	IN3	6931	40	9	27	1538	4636931	13	40
9 28	EA7	0500	20	9	28	1145	3270500	4	20
10 1	NE1	6592	15	10	1	1727	6316592	6	15
10 2	DI8	7506	20	10	2	1729	3487506	1	20
10 3	EA7	7883	30	10	3	1048	3277883	8	30
10 4	CE6	2411	20	10	4	1125	2362411	2	20
10 9	NE1	6592	10	10	9	952	6316592	5	10

4 2029	MESSAGE TOLL CHARGES	1 25	1 70	CALLING STA.	4 2029	TOTAL	2 95
FEDERAL TAX APPLICABLE							
TOTAL MESSAGE TOLL CHARGES							
TOTAL FEDERAL TAX							
TOTAL CARRIED TO BILL							

Figure 8. Block diagram of punched card tape-to-bill process

processing. This routine is to a large extent a manual procedure, consisting primarily of computing the cost of each call from the ticket data, adding federal (and sometimes local) taxes, and so forth. These calls are then transcribed to and added up on a toll service statement which is attached and added to the regular monthly bill. The amount of detail shown for each call is usually prescribed by the state or federal regulatory commissions under whose jurisdiction the telephone company operates.

Some of these regulatory bodies require all toll calls to be billed in more or less detail, while some permit calls within certain metropolitan areas to be reduced to message units. All message units are then added and bulk-billed for each billing period as one item without detail. In detail billing each call may show not only the called station, the date and total cost, but may also show rate and conversation time to permit subscribers to check each call in detail for themselves.

No matter what rules govern the manual toll billing process of an operating company, it is always a function which requires considerable organization details and manual labor. The use of various labor-saving devices has assisted in making manual accounting very efficient, but complete mechanization of this process will considerably reduce billing costs to telephone companies.

As described in previous sections, each toll call completed in a SATT system central office is recorded in a perforated tape. This tape is the equivalent of the call tickets obtained in manual operation and is therefore the primary record of each call.

Completed tapes from central offices are sent to an accounting center by mail or by other means. However, when trunk facilities are available, the tape data are also suitable for transmission by wire to an accounting center. This can in many cases speed up processing operations.

To obtain subscribers' toll statements from ticket tapes, two methods are available: manual or automatic bill processing.

Manual Processing

When it is desired to obtain subscribers' bills from SATT tapes by manual methods, a tape-to-ticket assembly is provided; see Figure 6. This assembly consists of a printer controlled by a tape reader. This assembly prints a standard size toll ticket for each call contained in the tape. These tickets are then processed in the same manner as regular manual tickets.

If desired, this process can be made partly automatic. A unit termed a computer can be included with the tape reader. When this is done, the cost of each call will be automatically computed from the tape data and printed on the ticket in addition to other data.

As previously described, the tabulator can be strapped in such a manner as to record each item in the tape in the same sequence as used on tickets prepared by manual operators. This will avoid manual errors in processing tickets from two sources.

Automatic Processing

Automatic accounting (processing) of SATT system call data is accomplished by means of punched card accounting equipment. This type of equipment has proved satisfactory over many years for almost every known kind of accounting requirement.

It may be of interest at this point to relate that surveys of accounting methods by independent telephone companies have disclosed the fact that some of them are already using punched card equipments to speed up and simplify manual accounting processes. When these companies adopt SATT operation, they will therefore be able to continue using their present facilities and methods by merely adding a tape-to-card unit to each card punch.

The first and basic step in obtaining printed subscribers' bills and auditing records with punched card equipment is to produce a punched card from the call data in a ticket tape. This is done by using a tape-to-card unit with a regular card punch, Figure 7. After a punched card has been obtained, it is used to control subsequent processes from which the basic end result will be a printed subscriber's bill, Figure 8. However, in this process, various other types of records are also obtained, depending on the amount and kinds of accounting data an operating company may require.

An interesting by-product obtainable from punched cards is toll traffic data. Computation machines can be set to sort out the particular items desired such as busy hour calls, average holding times, and from what classes of subscribers calls originate.

Conclusion

When the SATT system was designed, one of the primary aims was to use as far as possible tried and proved circuitry arrangements, and only such components as had been proved satisfactory in other fields and were commercially available. Thus the system described herein contains common control (Director) circuitry which has been used for many years in metropolitan areas. The record tape is of the same type as the message tape used in telegraph services. The punched card accounting equipment is also a well-known and proved equipment used in many commercial fields.

SATT system facilities are basically of such a nature that they can also be provided for non-Director step-by-step automatic central offices of all types and sizes, but special numbering and dialing methods must then be employed for subscriber dialed toll traffic. However, the end result is the same in each office: each completed subscriber dialed toll call is recorded in a ticket tape in exactly the same manner as in a Director-controlled office.

One interesting fact is already becoming apparent. Operating companies now providing SATT system service have found that mechanical accounting is not only a very desirable method as compared with manual processing, but also more economical.[6]

As automatic telephone switching eventually became the accepted mode in providing local telephone service, subscriber dialing, automatic call recording, and automatic accounting will without a doubt be the accepted mode in providing regional and national toll telephone service in the future.

References

1. Automatic Switching of Toll Traffic, **Dr. Arthur Bessey Smith.** *Journal*, Western Society of Engineers (Chicago, Ill.), volume 29, number 2, February 1924.

2. The Post Office and Automatic Telephones, **Col. T. F. Purves.** *Journal*, Institution of Electrical Engineers (London, England), volume 63, number 343, July 1925.

3. La Louviere Automatic Network of the Belgian Telephone System, **R. F. Stehlik.** *Automatic Electric Technical Journal* (Chicago, Ill.), volume 2, number 4, January 1951.

4. Application of Strowger Automatic Toll Ticketing and Punched Card Accounting in Metropolitan Networks, **J. T. Naylor.** *Automatic Electric Technical Journal* (Chicago, Ill.), volume 2, number 2, February 1950.

5. Automatic Switching for Nation-Wide Telephone Service, **A. B. Clark, H. S. Osborne.** *AIEE Transactions*, volume 71, part I, September 1952, pages 245–48.

6. Automatic Toll Ticketing in Accounting, **H. A. Hansen.** *Telephony* (Chicago, Ill.), volume 141, number 17, October 27, 1951.

No Discussion

Analysis of a Comb Filter Using Synchronously Commutated Capacitors

W. R. LePAGE
MEMBER AIEE

C. R. CAHN
ASSOCIATE MEMBER AIEE

J. S. BROWN
NONMEMBER AIEE

THE possibility of constructing a comb filter by commutating the input signal to a bank of capacitors through a series resistor has been suggested in the literature.[1] A theoretical investigation of the properties of such a filter is given in the following analysis.

The general scheme of the filter is shown in Figure 1. Rotating brush *A* is fed by the input signal (a zero impedance potential source) through a resistor *R* in series. While *A* is in contact with any one segment, the capacitor begins to change its charge, so that its potential begins to approach that of the input signal at that instant. However, the time constant *RC* is much longer than the time of dwell of the brush on one segment, so that the capacitor potential cannot change much during one duration of contact. If the signal frequency is a multiple of the frequency of rotation, the signal will have the same value each time the brush is in contact with a given segment, so that after many revolutions the potential across the capacitor can finally attain the value of the signal at the instant when the brush touches the segment. Random signals, aperiodic signals, and signals which are periodic but not of a frequency which is a multiple of the rotational frequency will not benefit from this storage process. Therefore, if the commutator is then scanned by an infinite impedance contact brush *B*, it will pick out a signal in which periodic components of the input signal are strongly predominant. In order not to affect the wave shape appreciably, it is necessary to have many commutator segments which will approximate a continuous distribution of segments.

One use of a comb filter is to improve the signal-to-noise ratio of a periodic signal mixed with noise. In the analysis to follow, the filter is regarded both as a frequency selective device and as a device which averages the noise over a large number of cycles of the periodic signal. The analysis shows that the signal-to-noise ratio may be improved approximately by the square root of the number of cycles over which the averaging extends, a result equivalent to that obtained by Harrington and Rogers.[2]

Analysis for Steady-State Response to a Sinusoidal Signal

The theoretical analysis is approximate, being based on the assumption that the input signal is "quantized" by resolving it into a staircase approximation, as shown in Figure 2(A). For convenience in the analysis, assume that the input signal is switched on at $t=0$, the instant that brush *A* is in the center of segment *O*, and that all capacitors are initially uncharged. The following is a partial list of notation:

f = input signal frequency
f_r = frequency of rotation of the brushes
N = number of capacitors and commutator segments
T_i = time required for a brush to pass from the center of one segment to the center of the next segment
T_d = time of dwell of a brush on one commutator segment
k = an index number for counting revolutions of the brushes
p = an index number for indicating capacitor position
$e(t)$ = the input signal as a function of time
$e_p(k)$ = potential difference of capacitor p after brush A has rotated past it k times

The analysis begins by considering the transient process by which a certain capacitor p changes its charge during the k'th revolution of brush *A*. Consider p fixed for successive values of k. For the general k'th revolution t lies in the time interval

$$\left(p+kN-\frac{1}{2}\right)T_i \le t \le \left(p+kN+\frac{1}{2}\right)T_i \qquad (1)$$

during which time interval the quantized input signal remains constant. For the transient state under consideration the potential difference of the capacitor is $e_p(k-1)$, at the beginning of the k'th contact of segment p with the brush. Also, during the interval (1) the quantized input signal is constant at the value $e[(p+kN)T_i]$. The charging circuit is shown in Figure 2 (B) for which the transient analysis gives

$$e_p(k)=(1-h)e[(p+kN)T_i]+he_p(k-1)$$

where

$$h=\epsilon^{-T_d/RC} \qquad (2)$$

Similar equations can be written for $e_p(k-1)$, $e_p(k-2)$, . . . and then substituted into each other to give the terminating series

$$\begin{aligned}e_p(k)&=(1-h)\{e[(p+kN)T_i]+ he[(p+kN-N)T_i]+ h^2e[(p+kN-2N)T_i]+\ldots+h^ke[pT_i]\}\\ &=(1-h)\sum_{r=0}^{k} h^r e[(p+kN-rN)T_i] \qquad (3)\end{aligned}$$

To obtain the steady-state response to a sinusoidal input, assume that the input signal is a sine wave, to be written

$$e(t)=E \sin 2\pi ft=EIm\epsilon^{j2\pi ft} \qquad (4)$$

where *Im* stands for the "imaginary part of." The intermediate steps are simplified by writing

$$a=\frac{2\pi f}{f_r}=2\pi NfT_i; \quad w=\frac{p}{N} \qquad (5)$$

so that $2\pi f(p+kN-rN)T_i=a(w+k-r)$. Then

$$e[(p+kN-rN)T_i]=EIm\epsilon^{ja(w+k-r)} \qquad (6)$$

Equation 3 can now be written

$$\begin{aligned}e_p(k)&=(1-h)EIm\epsilon^{ja(w+k)}\sum_{r=0}^{k} h^r\epsilon^{-jar}\\ &=(1-h)EIm\epsilon^{ja(w+k)}\times \left[\frac{1-h^{k+1}\epsilon^{-ja(k+1)}}{1-h\epsilon^{-ja}}\right] \qquad (7)\end{aligned}$$

by noting that the sum on the left is a geometric series of ratio $h\epsilon^{-ja}$. Equation 7 gives the distribution of potential around the commutator after each revolution of the brush.

We are primarily interested in the distribution of potential after many revolutions. Therefore, we note that as k becomes large the term h^{k+1} approaches zero, so that ultimately we can write

$$\begin{aligned}e_p(k)&=ImE_O\epsilon^{ja(w+k)}\\ &=E_O \sin [2\pi f(p+kN)T_i+\theta] \qquad (8)\end{aligned}$$

where E_O is the complex quantity

$$E_O=E_O\epsilon^{j\theta}=E\left(\frac{1-h}{1-h\epsilon^{-ja}}\right) \qquad (9)$$

Equations 8 and 9 embody the essential characteristics of the steady-state response and so bear further analysis. In practical cases h may be only slightly less than 1, in which cases a very large value of k must be attained before equation 8 can be used. In any case, equation

---

Paper 53-61, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted May 21, 1952; made available for printing December 3, 1952.

W. R. LePage, C. R. Cahn, and J. S. Brown are all with Syracuse University, Syracuse, N. Y.

This paper reports on work done under contract with the Rome Air Development Center of the United States Air Force.

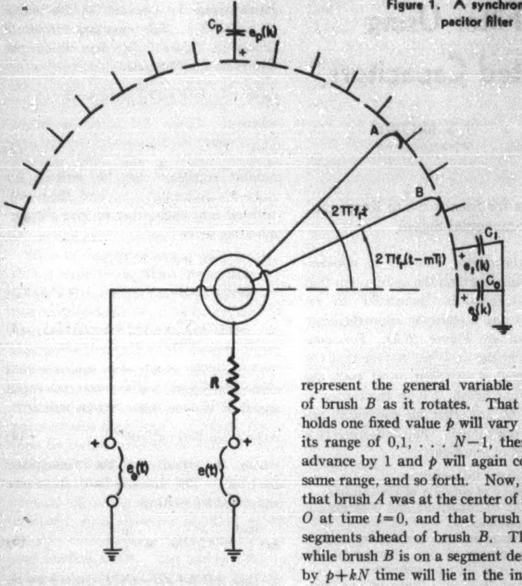


Figure 1. A synchronous capacitor filter

8 is the steady-state response and an h near unity means that the filter has a long build-up time.

Next, we are interested in discovering what sort of a signal brush B will pick up as it sweeps over the segments without drawing current. Equation 8 provides this information if $p+kN$ is allowed to represent the general variable position of brush B as it rotates. That is, as k holds one fixed value p will vary through its range of 0,1, . . . $N-1$, then k will advance by 1 and p will again cover the same range, and so forth. Now, we note that brush A was at the center of segment O at time $t=0$, and that brush A is m segments ahead of brush B. Therefore, while brush B is on a segment designated by $p+kN$ time will lie in the interval

$$(p+kN+m)T_i-\frac{T_d}{2}\leq t\leq(p+kN+m)+\frac{T_d}{2} \tag{10}$$

since brush A is on the segment designated by $p+kN+m$. At this point there is no longer a need for counting time intervals in terms of segment positions p and number of rotations k. Accordingly, we can let

$$n=p+kN+m \tag{11}$$

where n is the total number of segments passed over up to the time when t lies in the interval 10, not counting the starting position. Now, from equation 8 we obtain the simplified result

$$e_O'(t)=E_O\sin[2\pi f(n-m)T_i+\theta]$$

when

$$nT_i-\frac{T_d}{2}\leq t\leq nT_i+\frac{T_d}{2} \tag{12}$$

By allowing n to take on successive integral values, t progresses continuously, and equation 12 specifies a discontinuous output function which remains undefined for those time intervals when brush B is not on a segment. In a practical device the signal may be clamped at the value of the output for the immediately previous segment, or it may be reduced to zero by a blanking circuit. A typical graph of equation 12 is shown in Figure 3. In practice, the output can be smoothed by a low pass filter to yield an approximate sinusoid. For example, let $e_O(t)$ be the sinusoid

$$e_O(t)=E_O\sin(2\pi ft+\theta')$$

where

$$\theta'=\theta-2\pi fmT_i \tag{13}$$

It is noted that $e_O(nT_i)=e_O'(nT_i)$, and so it is concluded that equation 13 is a sinusoid passing through the mid-points of the steps of equation 13, and is approximately the smoothed output function.

Appraisal of the Response Function

As long as we remain within the frequency limit that allows the output function to be considered a sinusoid, the further study of the filter can be reduced to a consideration of Figure 4, where the response function $F(a)$ is obtained from equation 9 as

$$F(a)=\frac{1-h}{1-he^{-ja}} \tag{14}$$

which is a periodic function of the fre-

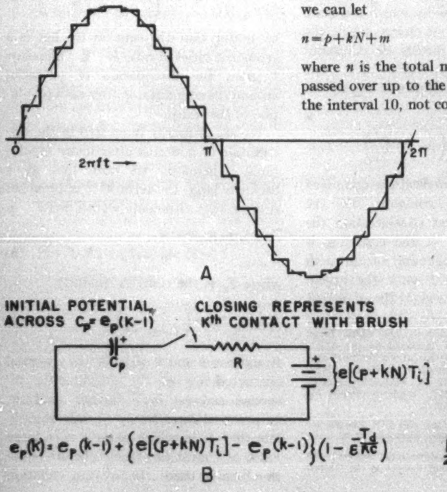
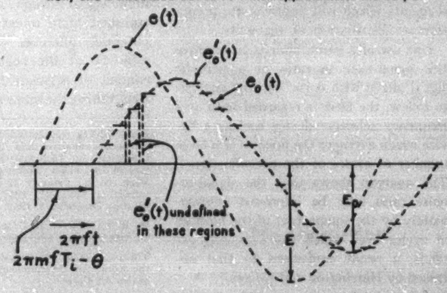

0
$2\pi ft\rightarrow$
π
2π
A
INITIAL POTENTIAL ACROSS $C_p = e_p(k-1)$
CLOSING REPRESENTS K^{th} CONTACT WITH BRUSH
C_p
R
+
$e[(p+kN)T_i]$
$e_p(k)=e_p(k-1)+\{e[(p+kN)T_i]-e_p(k-1)\}(1-\varepsilon^{-\frac{T_d}{RC}})$
B

Figure 2(A) (left). Approximation of a sinusoid by a staircase function. (B). The elementary charging circuit

Figure 3 (below). Output signal, showing actual discontinuous function, and a smoothed-out curve which approximates the output



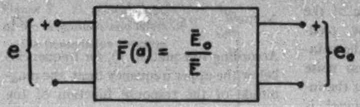


Figure 4. Definition of terms for treating the filter as a transmission system

$e = Im\bar{E}\epsilon^{j\omega t}$, where $\bar{E} = E\underline{/0}$ degrees

$e_0 = Im_r\bar{E}_0^{j\omega(t-mT_1)}$, where

$\bar{E}_0 = E_0\underline{/\theta}$

quency parameter a. Note that the time delay due to the separation of m segments between the two brushes has been excluded from $\bar{F}(a)$. A routine procedure shows that equation 14 can be written

$$F(a)=\frac{1}{\sqrt{1+\frac{2h}{(1-h)^2}(1-\cos a)}}\underline{/\theta} \tag{15}$$

where

$$\theta=-\tan^{-1}\frac{h\sin a}{1-h\cos a}$$

Equation 15 exhibits a maximum response whenever a is an integral multiple of 2π. The general form of the response function is shown in Figure 5.

By the analysis given in the Appendix, it is found that the bandwidth between the half-power points of any one of the response peaks is

$$B=\frac{T_d}{\pi RC}f_r=\frac{T_d/T_1}{\pi NRC} \tag{16}$$

It is recalled that the formulation of the response characteristics of the network represented by equation 15 is valid only after a sufficient number of revolutions to make the term h^{k+1} negligible compared with 1, in equation 7. Let us return now to equation 7 and define the response function

$$F(a,k)=\frac{(1-h)[1-h^{k+1}\epsilon^{-ja(k+1)}]}{1-h^{-ja}} \tag{17}$$

which approaches equation 14 as k becomes large. We are primarily interested in the build-up time at an integral multiple of f_r, for which a is an integral multiple of 2π. In that case equation 17 reduces to

$$1-h^{k+1}=1-\epsilon^{-(k+1)T_d/RC} \tag{18}$$

The response has risen to within $(1/\epsilon)^2$ of the maximum response when the exponent of ϵ on the right of equation 18 is -2, which gives

$$k_b=\frac{2RC}{T_d}-1\approx\frac{2RC}{T_d} \tag{19}$$

as the necessary number of revolutions. The corresponding build-up time t_b is

$$t_b=\frac{2RC}{T_d}\left(\frac{1}{f_r}\right)=\frac{2}{\pi B} \tag{20}$$

which shows an understandable relationship between build-up time and bandwidth.

The response function given by equation 14 can be further analyzed by writing

$$\bar{F}(a)=\frac{1-h}{1-h\epsilon^{-ja}}=\frac{1-h}{2}\left[1+\frac{1+h\epsilon^{-ja}}{1-h\epsilon^{-ja}}\right] \tag{21}$$

and noting that

$$\bar{G}(a)=\frac{1+h\epsilon^{-ja}}{1-h\epsilon^{-ja}} \tag{22}$$

is similar in form to the equation for the impedance of a transmission line with a real reflection coefficient h. Therefore, the known geometric properties of a transmission line impedance chart can be used to interpret equation 22 as a function of a.[3]

A typical plot of $\bar{F}(a)$ is given in Figure 6 for $h=7/8$. Much larger values of h will probably be used in practice. Loci of constant h are circles with centers on the real axis, and loci of constant a are circles centered on the line perpendicular to the real axis at $(1-h)/2$. The dimensions of the circles and locations of the centers are indicated in Figure 6.

As the frequency varies over the spectrum, the point of operation travels repeatedly around the circle of constant h.

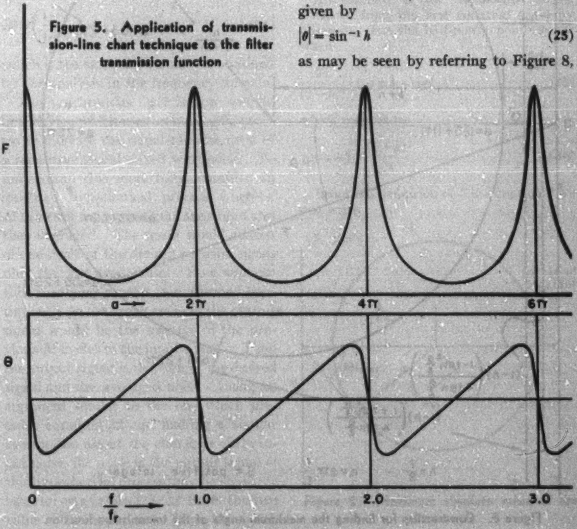


Figure 5. Application of transmission-line chart technique to the filter transmission function

This fact is indicated by labeling values of a with the additive term $2s\pi$, where s is a positive integer. The number of times the chart may be encircled depends upon the upper frequency limit of validity of the analysis, which in turn depends upon the number of segments. A plot of one peak of the response function and its associated angle function for the case $h=0.99$ is shown in Figure 7.

Approximate relations for the response properties can be given for certain important limiting cases. When h is approximately 1, the important part of the response curve occurs for small deviations of a from multiples of 2π. If such a deviation is denoted by δa, we can use the approximate relations $\cos a\approx 1-(\delta a)^2/2$ and $\sin a\approx\delta a$ in equation 15 to give

$$F(a)=\frac{1}{\sqrt{1+\frac{h(\delta a)^2}{(1-h)^2}}}\underline{/\theta} \tag{23}$$

where

$$\theta=-\tan^{-1}\frac{1}{\frac{1-h}{(\delta a)h}+\frac{\delta a}{2}}$$

Continuing to assume h close to unity, for large derivations of a from an integral multiple of 2π the first equation of 15 becomes

$$F(a)\frac{1}{\frac{2\sqrt{h}}{1-h}\sin\left(\frac{\delta a}{2}\right)}\underline{/\theta} \tag{24}$$

where θ continues to be given by the second equation of 15. It is interesting to note that $|\theta|$ has a maximum value given by

$$|\theta|=\sin^{-1}h \tag{25}$$

as may be seen by referring to Figure 8, which is derived from Figure 6. Thus, the maximum absolute value of the phase angle is very nearly 90 degrees.

It should be noted that superposition holds for the synchronous filter, as a result of the linearity of equation 3. Therefore, by using the Fourier series or integral in the usual way, the transmission properties for a sinusoidal wave can be extended to include nonsinusoidal signals. However, in this connection, it is important to realize that the filter cannot transmit a periodic function of frequency higher than $Nf_r/2$, and that as this limiting frequency is approached the response loses its accuracy.

Noise-Reducing Properties of the Synchronous Comb Filter, Analysis Using the Frequency Spectrum of Noise

One use of the synchronous filter is to increase the signal-to-noise ratio of a periodic signal mixed with noise. Because the response characteristic of the synchronous filter has narrow passbands about each harmonic of the rotational frequency f_r, there is less noise in the output of the filter than in the input. However, the periodic signal is passed with substantially no attenuation if its frequency is a multiple of f_r, provided that the filter has been operating for a time longer than t_b. If the average power per cycle of the input noise is given by the spectrum function $N(f)$ and the response characteristic of the filter has a magnitude $T(f)$, the average power (or mean square value) of the output noise over a bandwidth W is given by

$$\sigma_0^2 = \int^{W} N(f)|T(f)|^2 df \qquad (26)$$

assuming that W is less than the upper frequency limit of the filter. A quantitative estimate of the noise-reducing properties of the filter is made by assuming that the noise spectrum is flat with frequency in the range 0 to mf_r and is zero outside, where m is a positive integer such that mf_r is below the upper frequency limit of the filter. If the average power of the input noise is σ^2

$$N(f) = \frac{\sigma^2}{mf_r}, \quad 0 \leqq f \leqq mf_r \qquad (27)$$

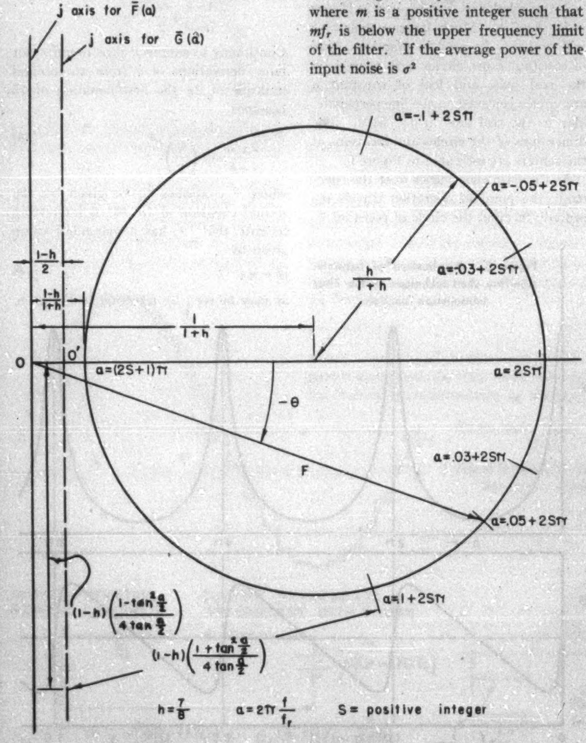


Figure 6. Construction for finding the maximum angle of the transmission function

According to equation 15, for frequencies below the upper frequency limit, the magnitude of the response function of the filter as a function of frequency is

$$T(f) = \frac{1-h}{\sqrt{1+h^2-2h\cos\left(\frac{2\pi f}{f_r}\right)}} \qquad (28)$$

If equations 27 and 28 are substituted into equation 26, subsequent simplification yields

$$\sigma_0^2 = \sigma^2\left(\frac{1-h}{1+h}\right) = \sigma^2\left[\frac{1-\epsilon^{-T_d/RC}}{1+\epsilon^{-T_d/RC}}\right]$$
$$= \sigma^2\left[\tanh\left(\frac{T_d}{2RC}\right)\right] \qquad (29)$$

Thus, the rms noise reduction factor is

$$\frac{\sigma}{\sigma_0} = \frac{1}{\sqrt{\tanh\left(\frac{T_d}{2RC}\right)}} \approx \sqrt{\frac{2RC}{T_d}} \qquad (30)$$

since ordinarily $T_d << RC$. Using equation 20, equation 30 may be written

$$\frac{\sigma}{\sigma_0} \approx \sqrt{\frac{2f_r}{\pi B}} = \sqrt{t_b f_r} \qquad (31)$$

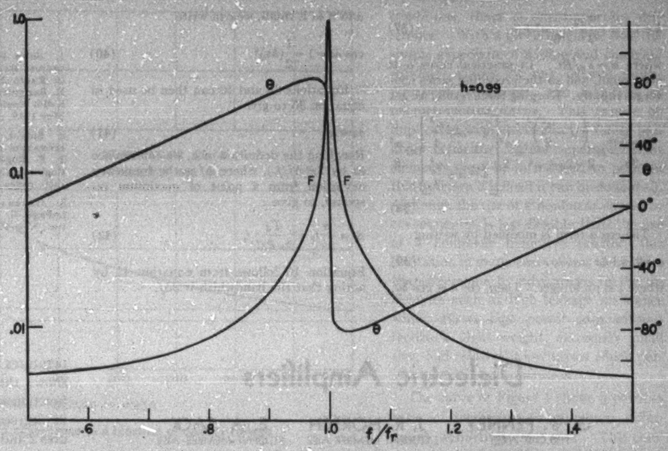
If the periodic signal has a frequency which is a multiple of f_r, σ/σ_0 is the ratio of the signal-to-noise ratio at the output divided by the signal-to-noise ratio at the input. Note that $t_b f_r$ is the build-up time measured in revolutions of the brush.

Noise with a nonflat spectrum will experience about the same reduction, provided that the noise spectrum varies slowly from peak to peak of the transmission function.

Analysis Using Statistical Properties of Noise in Time

An alternative point of view of the noise reduction process is obtained from an analysis in time rather than frequency. The assumption is continued that the highest frequency component of the noise is below the upper frequency limit of the synchronous filter. In the time domain this means that the noise cannot change significantly during the time of dwell of a brush on a segment. The assumption is also continued that the filter has been in operation for a time exceeding its build-up time. Since the operation of the filter is linear, the noise may be considered separately from the useful signal, so that it may be assumed that the input signal is a random noise function. The resulting potential difference across the p'th capacitor after k'th revolution is still given by equation 3. Since $e_p(k)$ is the sum of a number of random variables, it is itself a random variable.

Figure 7. A typical "cycle" of the magnitude and angle of the transmission function



Inspection of equation 3 reveals that the individual terms in the summation for $e_p(k)$ are samples of $e(t)$ taken at a sequence of instants of time separated by the period of rotation. This summation starts at the time of observation and runs back toward zero time. However, for practical purposes, the summation can be considered stopped at the value of r for which h^r is reduced to $1/\epsilon^2$. According to equation 18, such a value of r is also the number of revolutions required for a sinusoidal response to build up. Therefore, the conclusion is reached that for k greater than the build-up time in revolutions $e_p(k)$ will be substantially independent of the values of k and p. Recalling that the output signal obtained from brush B consists merely of samples of $e_p(k)$, it follows that the output noise can be considered as the result of a stationary random process. By the ergodic hypothesis, for a stationary random process the mean square value considered as an average in time equals the mean square value considered as an ensemble average. This latter value may be obtained by making the assumption that the samples of $e(t)$ mentioned previously are statistically independent. This is accurately true for "white" noise and approximately true for noise which varies slowly from peak to peak of the transmission characteristic. By a fundamental theorem for the sum of a number of independent random variables

$$\overline{[e_p(k)]^2}=(1-h)^2\sum_{k=0}^{k}\overline{\{h^r e[(p+kN-rN)T_1]\}^2} \tag{32}$$

where the bar over a quantity means the ensemble average. However, the mean square value of the input noise, considered either as a time or ensemble average, is σ^2 so that equation 32 can be written

$$\overline{[e_p(k)]^2}=(1-h)^2\sum_{r=0}^{k} h^{2r}\sigma=\frac{(1-h)^2}{1-h^2}\times \sigma^2[1-h^{(2k+1)}] \tag{33}$$

Then, as k becomes very large, the left-hand side of equation (33) becomes σ_0^2, and the right-hand side simplifies to give the final result

$$\sigma_0^2=\left(\frac{1-h}{1+h}\right)\sigma^2 \tag{34}$$

which is the same answer as was obtained by the analysis in the frequency domain.

The synchronous filter is now seen to be a device which uses an averaging process to increase the signal-to-noise ratio of a repetitive signal mixed with noise. To understand this more fully, consider an idealized hypothetical process whereby M cycles of the input signal are stored and then averaged. The result would consist of one cycle of the desired periodic signal plus the averaged noise. Now suppose this averaging process is performed continuously so that each cycle of the output signal would be the average of the previous M cycles of the input signal. Then the output signal is the sum of the desired signal and the averaged noise. Using an argument similar to the one which precedes equation 32 and making a similar assumption about the character of the input noise, the noise in the output signal of this hypothetical process may be shown to have an rms values $1/\sqrt{M}$ times the rms value of the input noise. Thus, the signal-to-noise ratio at the output would be $\sqrt{M}$ times the signal-to-noise ratio at the input. This conclusion for the idealized process is interesting because it shows that this is the best result that can be obtained with an averaging process if the delay is limited to M cycles of the periodic signal. Essentially the same result was obtained for the actual device.

Appendix. Estimate of Bandwidth

An estimate of the bandwidth can be obtained from the first equation of 15 by observing that the half-power points occur when

$$\frac{2h}{(1-h)^2}(1-\cos a)=1 \tag{35}$$

which reduces to

$$\cos a=1-\frac{(1-h)^2}{2h} \tag{36}$$

From the definition of h in equation 2 we can write

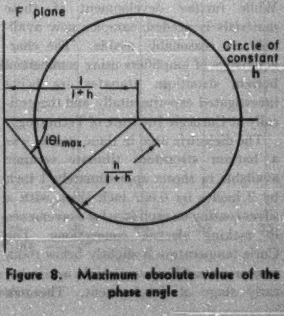


Figure 8. Maximum absolute value of the phase angle

$$h=1-\alpha+\frac{1}{2}\alpha^2-\frac{1}{6}\alpha^3+\ldots \qquad (37)$$

where $\alpha=T_d/RC$. In practical cases α is very small, and so the preceding series converges rapidly. Keeping three terms, we get

$$\frac{(1-h)^2}{2h}=\frac{\left(\alpha-\frac{\alpha^2}{2}\right)^2}{2-2\alpha+\alpha^2}=\frac{\alpha^2}{2}\left[\frac{1-\alpha+\frac{\alpha^2}{4}}{1-\alpha+\frac{\alpha^2}{2}}\right]\approx\frac{\alpha^2}{2} \qquad (38)$$

The cos a term is modified by writing

$$a=2\pi s+\delta a \qquad (39)$$

where s is an integer. Then, cos $a=$ cos δa, and if δa is small, we can write

$$\cos\delta a=1-\frac{1}{2}(\delta a)^2 \qquad (40)$$

Equations 38 and 40 can then be used in equation 36 to give

$$\delta a=\alpha \qquad (41)$$

Recalling the definition of a, we can replace δa by $2\pi(\Delta f/f_r)$, where Δf is the frequency deviation from a point of maximum response, to give

$$\Delta f=\frac{\alpha}{2\pi}f_r=\frac{T_d}{2\pi RC}f_r \qquad (42)$$

Equation 16 follows from equation 42 by noting that the bandwidth is $2\Delta f$.

References

1. Some Relations Between Speed of Indication, Bandwidth and Signal-to-Noise Ratio in Radio Navigation and Direction Finding, **H. Busignies, M. Dishal.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 37 May 1948, pages 478–88.

2. Signal-to-Noise Improvement Through Integration in a Storage Tube, **J. V. Harrington, T. F. Rogers.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 38, October 1950, pages 1197–1203.

3. General Network Analysis (book), **W. R. LePage, S. Seely.** McGraw-Hill Book Company, Inc., New York, N. Y., 1952, chapter 10.

No Discussion

Dielectric Amplifiers

G. W. PENNEY
FELLOW AIEE

J. R. HORSCH
STUDENT MEMBER AIEE

E. A. SACK
STUDENT MEMBER AIEE

Synopsis: Capacitors whose capacitance is a function of voltage can be made using commercially available barium strontium titanate ceramics. Such nonlinear capacitors can be used as controllable impedances to form a dielectric amplifier. Using nonresonant circuits without transformers or feedback, amplifiers have been tested and found to give a large power amplification, but with an output voltage less than the input voltage. Using resonant circuits a voltage amplification of four has been obtained with the same materials.

THE idea of a dielectric amplifier using nonlinear capacitors as the controllable impedance has been suggested before.[1,2] The development of the device has recently been stimulated by improved materials and by increased emphasis on more reliable electronic components for military equipment. High permittivity ceramics such as the barium strontium titanates provide dielectrics having both a high permittivity and a marked nonlinearity at ordinary temperatures. While further development of these materials is needed, samples now available are reasonably usable. The characteristics of amplifiers using commercial barium strontium titanates are being investigated experimentally and theoretically at Carnegie Institute of Technology.

The dielectric used in these studies was a barium strontium titanate ceramic available in sheets approximately 1 inch by 3 inches by 0.007 inch thick with a silver coating on each face for convenience in making electric connections. The Curie temperature is slightly below room temperature. Such materials are in an early stage of development. The one chosen seems to be the best that is available, but it is hoped that further development will result in dielectrics having higher electrical breakdown, lower losses, a greater nonlinearity, and especially more uniform properties. At present pieces cut from a given sheet show considerable variation.

The nonlinearity of the dielectric requires a modification of the usual definition of capacitance. In general one can use definitions of capacitance paralleling definitions of inductance. The characteristics of several barium strontium titanate capacitors have been determined by measuring current, voltage and frequency and calculating the capacitance. The capacitance measured varies with applied voltage. As the voltage is reduced to small values (1 to 5 volts) the capacitance becomes constant and approaches the differential capacitance dQ/dV.

Values of this differential capacitance have been determined as a function of d-c bias or control voltage for various temperatures and frequencies. Figure 1 shows the change in capacitance as a function of control voltage as measured at 25 degrees centigrade and 655 kc. Figures 2 and 4 show capacitance as a function of temperature and frequency. In Figure 4 at 300 kc there is a marked decrease in capacity. Since Von Hipple[3] has observed similar phenomena this has not as yet been investigated further. If capacitance is expressed as a percentage of that measured at zero control voltage and plotted as a function of control voltage, the variation with temperature and frequency appears less serious. This is shown by the curves of Figures 3 and 5.

If the experimental plot of differential capacitance versus control voltage is integrated with respect to voltage, the resulting curve may be used to better advantage for the analytical predetermination of dielectric amplifier characteristics. Such a plot of charge per unit area D as a function of direct potential gradient E is shown in Figure 6. A plot of D = 0.052 tanh 0.84E is shown for comparison.

The Nonresonant Amplifier in Steady State

Circuital applications of nonlinear reactors have been many and varied, especially during the last several years when refinement of the magnetic amplifier has greatly broadened its realm of usefulness. This section is a discussion of the use of nonlinear capacitors as variable reactive circuit elements, and as such it presents representative amplifier circuits, their general performance, and analytical methods for the predetermination of operating characteristics.

In order to establish a common basis of terminology, the term "dielectric amplifier" will be used in accordance with the following definition which is consistent

---

Paper 53-130, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted October 20, 1952; made available for printing December 8, 1952.

G. W. Penney, J. R. Horsch, and E. A. Sack are with Carnegie Institute of Technology, Pittsburgh, Pa.

This work has been done in part under Office of Naval Research Contract *N9-onr-30308*, and in part by J. R. Horsch and E. A. Sack in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Carnegie Institute of Technology.

The authors wish to acknowledge the comments and advice of Professors L. A. Finzi and E. M. Williams.

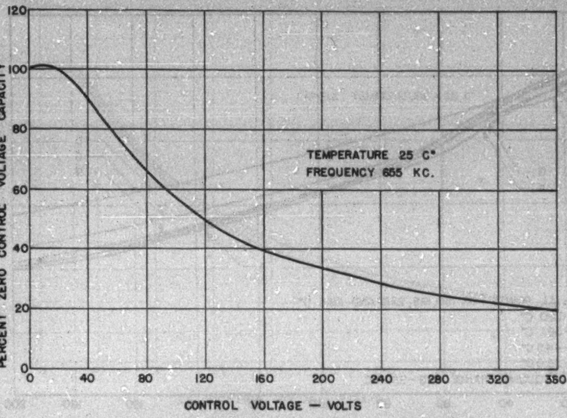


Figure 1. Change in capacitance as a function of control voltage

with general reactance amplifier terminology. "A dielectric amplifier is a device which in order to obtain amplification or control uses nonlinear capacitors either alone or in combination with other circuit elements."

An elementary circuit which exploits dielectric nonlinearity is that shown in Figure 7, but passage of load current by the input device is required. A better arrangement is to use blocking reactive elements to isolate the input and load circuits to prevent unwanted interaction. Figure 8 shows such a circuit. Capacitor *C*, which may also be made nonlinear for maximum gain performs the vital function of preventing voltage division of the biasing voltage by resistors R_P and R_L; this means that full bias is applied to the capacitor at all times.

The need for the blocking reactance L_B which prevents passage of supply frequency currents through the control circuit can be obviated by a further refinement shown in Figure 9. Here all voltages of fundamental frequency or odd harmonics thereof cancel from points 1 to 2; hence currents of these frequencies do not flow in the input circuit when the nonlinear capacitors are a matched pair. Thus only even harmonics of supply frequency flow through the input device. Since, as Figure 5 shows, these titanate elements exhibit a slowly varying ratio of *D* to *E*, they generate harmonics of small magnitude and consequently the input device of Figure 9 experiences little effect from the load circuit.

Superficial considerations seem to dictate that since a dielectric-magnetic analogy can be derived, it would be possible simply to transfer present information concerning magnetic amplifier circuits to those containing nonlinear dielectrics. Further investigation shows that while an analogy exists under certain conditions, there is one important difference. With a given magnetic core the simple expedient of adding and changing windings allows for any desired multiplications, additions, and isolation of magnetomotive forces. This cannot be duplicated with electromotive forces in a given capacitor; rather separate circuit elements must be introduced to perform this function if indeed it can be done at all. Although the use of a nonlinear capacitor consequently is less flexible than the use of a nonlinear inductive reactor, this limitation in many cases seems to be more than overcome by advantages of the dielectrics such as high leakage resistance which allows high power gain without feedback, light weight, extremely small size, and relatively low losses which permit tuned circuit operation.

The curve of Figure 1 shows a possible ratio of maximum to minimum differential permittivity of 5 to 1. This compares with a maximum usable ratio of differential permeabilities of about 25 to 1 for high-frequency ferromagnetic materials and 6,000 to 1 for alloys such as Deltamax. This comparatively small nonlinearity of dielectrics is a disadvantage

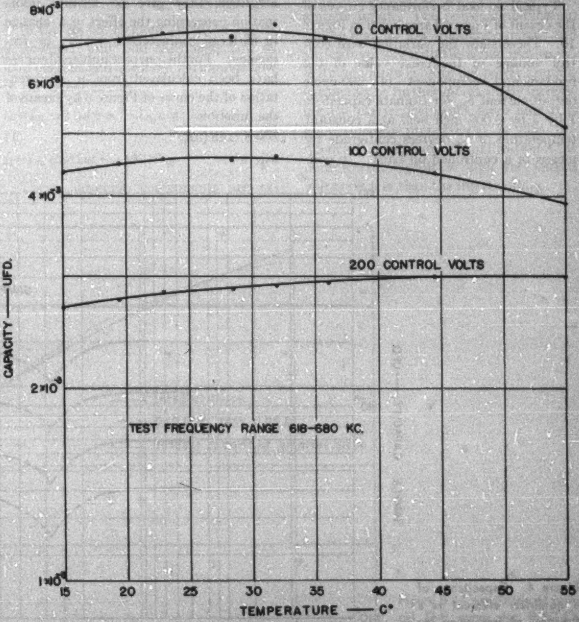
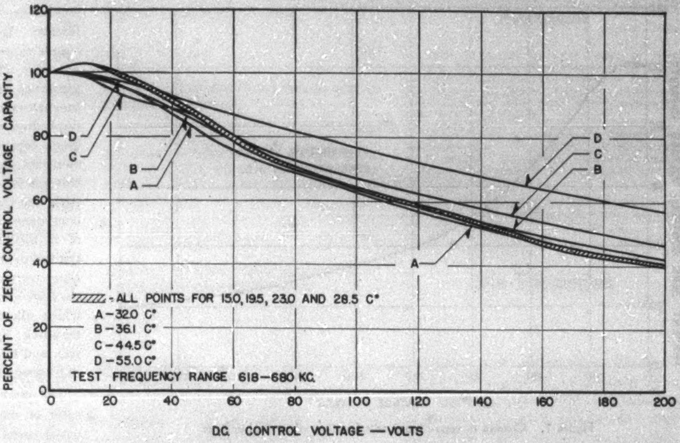


Figure 2. Capacitance of a nonlinear element as a function of temperature for various values of control voltage

Figure 3. Change in capacitance as a function of control voltage for various temperatures



from the standpoint of amplification but since harmonic generation is a direct function of nonlinearity, the harmonic content of a dielectric amplifier output is small as compared to that of magnetic amplifiers.[4]

Typical no-load performance curves of the circuit of Figure 9 are shown in Figure 10. These show the variation with control voltage of the crest value of the fundamental component of 500-cycle output current I_{m_1} for titanate capacitors 1 by 3 by 0.006 inch held at a constant temperature of 35 degrees centigrade by means of a controlled oil bath. In contrast with the control characteristics of a magnetic amplifier, dielectric amplifier output decreases as a state of saturation is approached.

Since the curves of Figure 10 show the performance of only one particular amplifier, they give no definite information concerning the effect of a change in material characteristics, size, or frequency. For this purpose universal curves have been calculated from a representation of the curve of Figure 6 by means of the function

$$D = A \tanh (aE) \qquad (1)$$

where

A and a are constants of the particular dielectric material under consideration, hence independent of capacitor size and connection

D is electric displacement, which for a parallel plate is equal to the charge per unit area on the plates

E is the gradient of the applied electromotive force

For a given capacitor of surface area S and thickness l

$$D = q/S \qquad (2)$$

$$E = v_N/l \qquad (3)$$

where q is the instantaneous charge on the capacitor and V_N is the applied voltage.

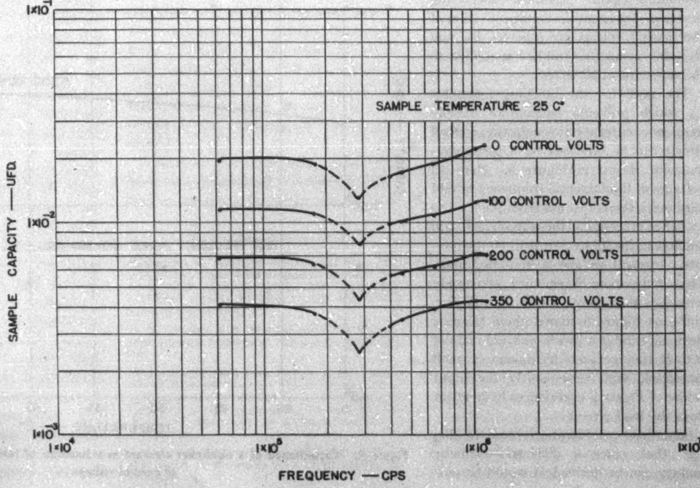


Figure 4. Capacitance of a nonlinear element as a function of frequency for various values of control voltage

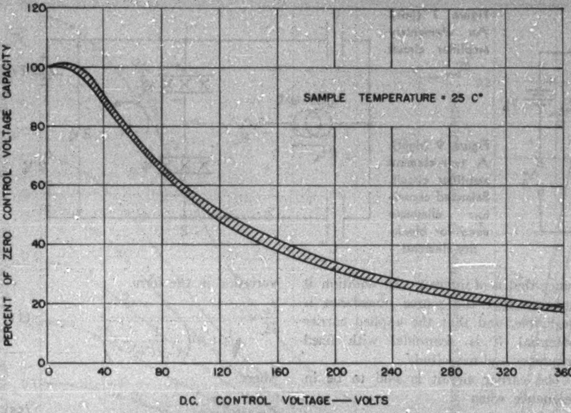


Figure 5. Change in capacitance as a function of control voltage. Range of values as measured at 61.7, 172.4, 286, 400, 455, 655, 1,025, and 1,166 kc

Thus

$$q/SA = \tanh(av_N/l) \tag{4}$$

In order to translate charge into current, q must be differentiated with respect to time. When V is harmonic and periodic, q is given by the Fourier series

$$q = Q_0 + Q_{m_1}\sin(\omega t+\varphi_1) + Q_{m_2}\sin(2\omega t+\varphi_2) + \ldots \tag{5}$$

Thus

$$i = \dot{Q}_0 + \omega Q_{m_1}\cos(\omega t+\varphi_1) + 2\omega Q_{m_2}\cos(2\omega t+\varphi) + \ldots \tag{6}$$

Since q/SA is dimensionless, $i/\omega SA$ is also dimensionless and on the universal curves is denoted as I_d.

In the single element circuit of Figure 8, the instantaneous value of load current contains both even and odd harmonics, so that

$$i_L = I_{m_1}\cos(\omega t+\varphi_1) + I_{m_2}\cos(2\omega t+\varphi_2) + I_{m_3}\cos(3\omega t+\varphi_3) \tag{7}$$

In the double element circuit of Figure 9, however, only odd harmonics flow as load current and the even harmonics flow in the signal circuit.

$$i_L = I_{m_1}\cos(\omega t+\varphi_1) + I_{m_3}\cos(3\omega t+\varphi_3) + \ldots \tag{8}$$

$$i_c = I_{m_2}\cos(2\omega t+\varphi_2) + \ldots \tag{9}$$

These harmonic components are expressed in general form in the following equations

$$\begin{aligned} I_{m_1} &= \omega SAI_{d_1} \\ I_{m_2} &= 2\omega SAI_{d_2} \\ I_{m_3} &= 3\omega SAI_{d_3} \end{aligned} \tag{10}$$

Higher harmonics are usually negligible. The dimensionless coefficients I_d are derived in Appendix 1 and plotted in Figures 11, 12, 13, and 14.

Johnson and Latson[6] have shown a graphical technique which determines the effect of resistive loading from no-load reactance amplifier curves. For small amounts of harmonic generation by the nonlinear device, a resistive load prescribes an elliptical load line. The dotted ellipse of Figure 9 serves for one particular value of load. The phase angle φ_1 is determined by the ratio of resistive to reactive voltage drops.

As a means of demonstrating a dielectric amplifier, the audio amplifier of Figure 15 was constructed. Figure 16 shows the signal volt-ampere input and power output as functions of signal frequency when a 50-kc carrier is used. The power factor of the signal circuit is estimated to be about 2 per cent so that the watts input is about 1/50th of the volt-ampere input. The power amplification can be high for low-frequency signals and varies inversely with the signal frequency. With adequate cooling the power amplification can be further increased by increasing the carrier frequency since output is approximately proportional to this frequency. While power amplification of the nonresonant circuit is high the output voltage is less than the input voltage.

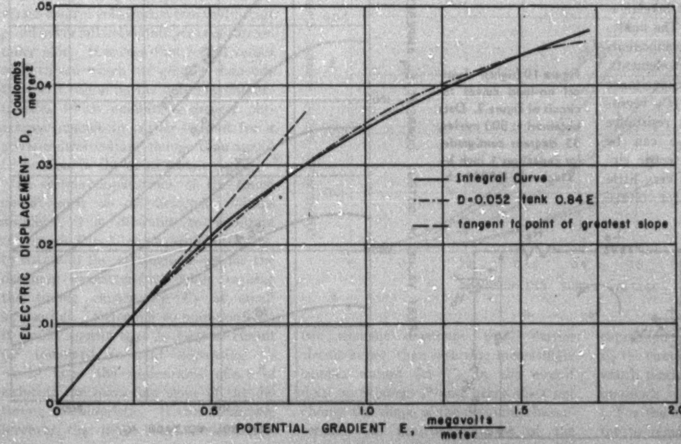


Figure 6. Comparison of typical D-E curve for $B_aS_rT_iO_3$ ceramic and D=0.052 tanh 0.84E

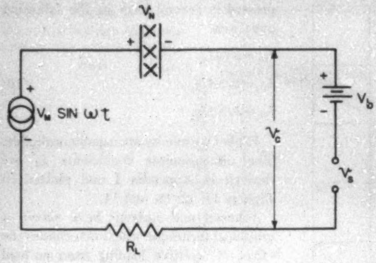


Figure 7 (left). An elementary amplifier circuit

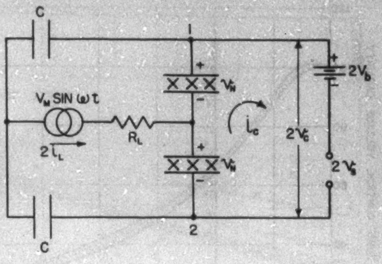


Figure 9 (right). A twin-element amplifier circuit. Balanced capacitors eliminate need for blocking elements

Series Resonant Amplifier

In the preceding section the carrier current control results from the change in impedance of the nonlinear capacitor as the control voltage varies. In such circuitry the carrier current can change no more, percentagewise, than the capacitance of the nonlinear element changes. It has been found that using the nonlinear capacitor as part of a tuned circuit permits a much greater degree of current control.

To investigate the series resonant circuit an analysis was chosen which is one compromise between rigor and convenience. This analysis considers the nonlinear element to have a capacitance whose magnitude cannot be changed by the carrier voltage appearing across it but can be varied by the control voltage. It is to be expected that although this procedure should permit a good approximation to the circuit response when the carrier voltages across the element remain small, it can serve only as a guide to experimental work when larger voltages are encountered.

The basic circuit for this investigation is shown in Figure 17. The analysis requires that all circuit components except C_N and R_N be linear elements. It assumes that a differential capacitance versus control voltage curve (C_N versus V_c) and an equivalent loss resistance versus control voltage curve can be defined for the nonlinear dielectric element but that the latter shows very little variation in R_N over the control voltage range that is of interest. In addition it assumes that the source impedance is negligible, and that the applied carrier potential V is sinusoidal with fixed frequency and magnitude.

The carrier circuit is said to be in resonance when

$$\omega L_T = \frac{1}{\omega\left(\frac{C_{N_r}C_T}{C_{N_r}+C_T}\right)}$$

Subscript r denotes circuit conditions at this point.

It is desired to compare the percentage variation of carrier current with control voltage in the tuned circuit with that variation which could be obtained in the nonresonant circuit. By writing *a-c* circuit theory equations for the carrier loop it is possible to express this variation in the form

$$\frac{I_L}{I_{L_r}} = \frac{1}{\sqrt{1+P^2\left(\frac{1}{C_N/C_{N_r}}-1\right)^2}} \tag{11}$$

where

$$P = \frac{QF}{\left(\frac{C_{N_r}}{C_T}+1\right)} \tag{55}$$

$$QF = \frac{\omega L_T}{R_N+R_{cT}+R_{LT}} \tag{40}$$

The derivation of this relation is shown in Appendix II, section 1. QF for the dielectric element is closely analogous to the familiar "quality factor" of linear circuitry so long as the carrier voltage across the nonlinear element remains small.

A family of curves showing I_L/I_{L_r}

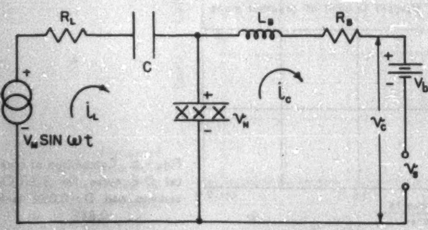


Figure 8 (below). A single element amplifier circuit. L_B and C used for isolation of control and supply

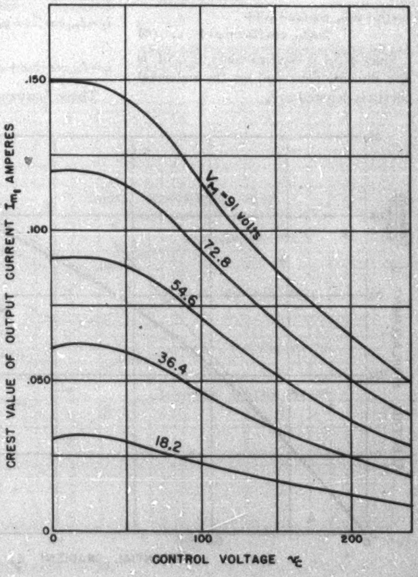


Figure 10 (right). Typical no-load curves for circuit of Figure 7. Data obtained at 500 cycles, 35 degrees centigrade, for capacitors 1 inch by 3 inches by 0.006 inch

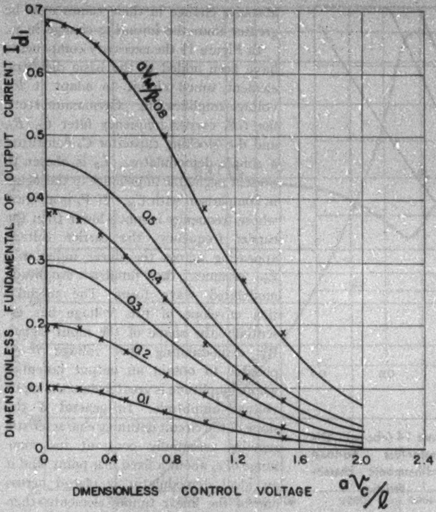


Figure 11. Dimensionless no-load control characteristics. Curves derived from $q/SA = \tanh(aV_a/l)$. Experimental points from Figure 8 shown by x's

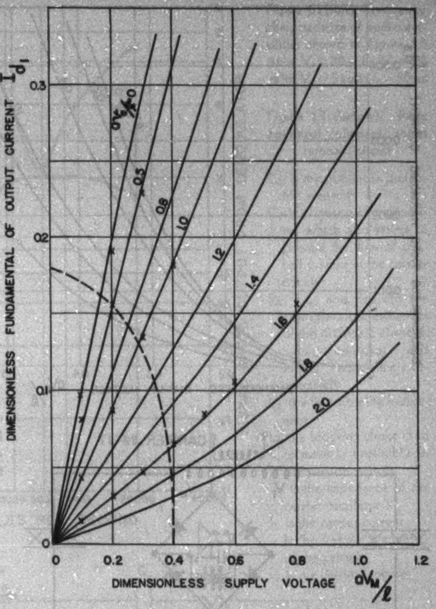
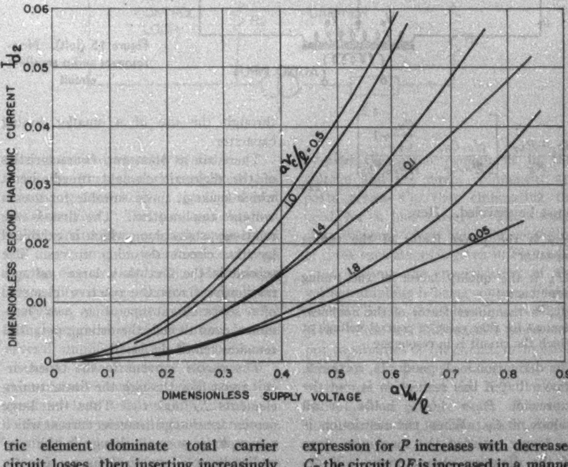


Figure 12 (above). Cross plot of Figure 11 to give dimensionless no-load output characteristics. Dotted ellipse used to determine effect of resistive loading by method of load lines

Figure 13 (below). Dimensionless no-load second-harmonic characteristics



versus C_N/C_{N_r} can be constructed for various values of P. Elimination of C_N/C_{N_r} between these curves and the C_N/C_{N_r} versus v_c curve gives a family of "detuning characteristics" which shows I_L/I_{L_r} versus v_c. Figure 18 compares detuning characteristics for various values of P with the experimental C_N/C_{N_r} versus v_c plot on which they are based. If the dielectric element were used in the nonresonant circuit the per cent current versus control voltage curve for that circuit could never fall below this per cent capacitance plot. It is seen that for all values of P shown which are greater than one there is a region on the detuning characteristics which exhibits a greater percentage change in carrier current for a given control voltage change than would be possible in the nonresonant circuit.

In certain applications of the tuned circuit such as a dielectric voltage amplifier, it is desirable to maintain the circuit P as large as possible (that is, to maintain the maximum slope for the detuning characteristic) while making the tuning capacitance C_T as small as possible. Referring to equation (55) it would appear that so long as circuit QF remained constant decreasing C_T would have the undesirable effect of reducing the maximum slope of the detuning characteristic. It can be shown, however, that if the losses in the dielectric element dominate total carrier circuit losses, then inserting increasingly smaller values for C_T in the circuit does not change P and hence does not change the shape of the detuning characteristic. As the denominator of the expression for P increases with decreased C_T the circuit QF is increased in a manner which tends to offset any change in the quotient.

For the losses in the nonlinear capacitor to remain dominant the restriction

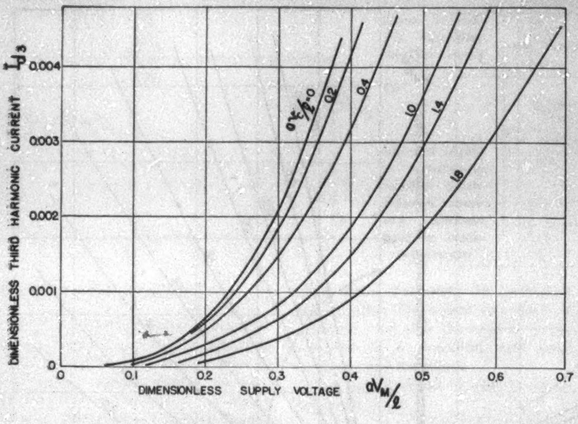


Figure 14 (above). Dimensionless no-load third-harmonic characteristics

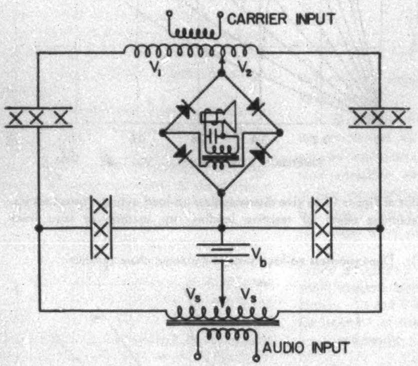


Figure 15 (left). Nonresonant audio amplifier circuit

$$\frac{R_{C_T}+R_{L_T}}{R_N}=\left[\frac{pfc_T C_{N_r}}{pf_{N_r}C_T}+\frac{\dfrac{C_{N_r}}{C_T}+1}{pf_{N_r}QF_L}\right]<<1 \qquad (12)$$

must be satisfied. Here

pfc_T is the power factor of the tuning capacitor

QF_L is the quality factor of the tuning inductance

pf_{N_r} is the power factor of the nonlinear element for that value of control voltage at which the circuit is in resonance

The derivation in Appendix II, section 2, shows that if this restriction is met the expression $P = 1/pf_{N_r}$ holds for all values of C_T. When the restriction is no longer satisfied the maximum slope of the detuning characteristic will decrease. The dielectric voltage amplifier is one application of the tuned circuit for which such a decrease is more than compensated by an advantage gained through the use of a smaller tuning capacitor.

There are at least two characteristics of the dielectric element tuned circuit which make it quite suitable for use in voltage amplification. The first is the relatively steep slope which is exhibited by the circuit detuning curves. The second is the fact that large voltages may exist across the reactive elements of a series circuit operating near resonance even though the driving potential remains relatively small.

The carrier current in the tuned circuit must flow through the linear tuning elements L_T and C_T. Thus the large percentage change in carrier current which accompanies a small change in the control voltage is matched by an equally large percentage change in the carrier voltages across the linear elements. Since the magnitudes of these voltages can be quite large to begin with, the absolute change in their values may be greater than the initiating change in v_c.

In Figure 19 the necessary components have been added to the basic dielectric element tuned circuit to adapt it for voltage amplification. Germanium rectifier *GR*, carrier frequency filter C_f, R_f, and d-c blocking capacitor C_0 constitute a simple demodulator. L_B is chosen to present negligible impedance to the signal in comparison with C_N. If V_c is a signal whose frequency is much lower than the carrier frequency, the carrier voltage appearing across the linear inductance L_T assumes the familiar amplitude modulated wave form. The modulation envelope of this voltage has essentially the shape of the input signal. By demodulating this voltage it is possible to obtain an output potential whose amplitude is greater than the input voltage amplitude. In general if the slope of the circuit detuning characteristic remains essentially constant for some range of v_c about a fixed bias point, and if an ideal demodulator is placed across one of the linear tuning elements, then for input signals whose amplitude does not exceed the prescribed range in v_c and whose frequency is small in comparison with the carrier frequency, $VA = 1.414 GV_{T_r}$. Here

VA is the ratio of the output signal voltage to the input signal voltage

G is the magnitude of the slope of the detuning characteristic

V_{T_r} is the rms value of the carrier voltage that appears across the linear tuning element at resonance

The derivation for this expression is shown in Appendix II, section 3.

In a preceding paragraph it was noted that inserting small values for C_T into the carrier circuit may not greatly reduce the slope of the detuning characteristic, and, indeed, if the losses in the dielectric element dominate circuit losses it will not change the slope at all. Figure 20 compares experimental characteristics for circuits having $C_T = 10\ C_{N_r}$, $C_T = C_{N_r}$, and $C_T = 0.1\ C_{N_r}$ with the analytical detuning characteristic for $P = 40$. These circuits were constructed using commercially available coils and capacitors. It can be seen that decreasing C_T to the point where $C_T = C_{N_r}$ makes only a negligible change in the slope of the detuning curve, and decreasing C_T to the point where $C_T = 0.1\ C_{N_r}$ reduces the slope by a factor of only 2.8. At the same time if the carrier potential across C_N is maintained at a fixed value the resonant carrier voltage across C_T and C_N in series and hence across L_T is increased as C_T is reduced. Herein lies the significant advantage of adding a

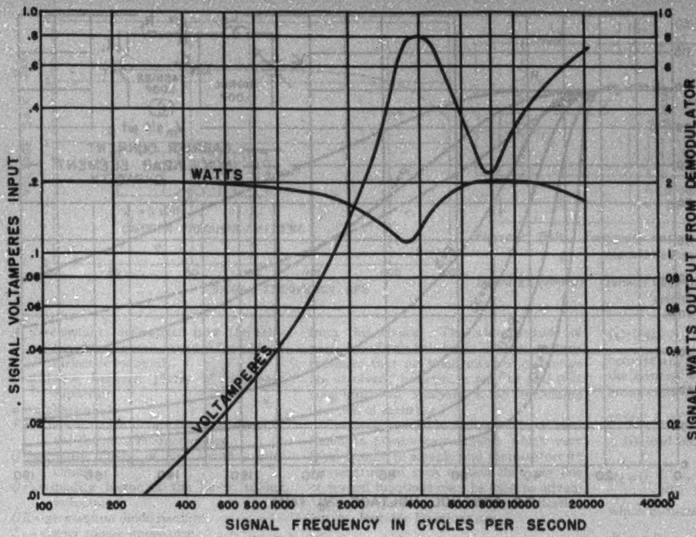
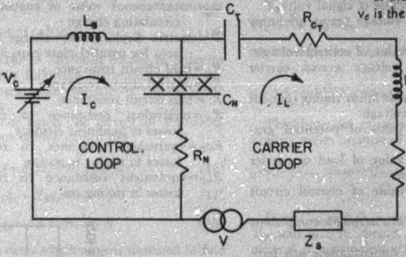


Figure 16 (left). Frequency characteristics of audio amplifier shown in Figure 15 when $V_1=30$ volts, $V_2=20$ volts, $V_b=5$ volts, $f_c=50$ kc

Figure 17 (below). Basic nonlinear element series resonant circuit

C_N is the nonlinear dielectric element
C_T is a linear tuning capacitor which also serves to prevent the flow of control current in the carrier circuit
R_N, R_{C_T}, and R_{L_T} are the equivalent loss resistances of the dielectric element, tuning capacitor, and tuning inductance respectively
L_T is a linear tuning inductance
L_B is a blocking choke used to make I_C negligible in comparison with I_L
Z_s is the impedance of the carrier oscillator V
I_L is the carrier current
ω is the angular frequency of the carrier
V_c is the control voltage



linear tuning capacitor to the circuit. For a given carrier potential across C_N the signal voltage can exert the same percentage control over a significantly larger voltage as C_T is decreased. This results in a greater absolute change in the carrier potential across L_T for a given signal voltage variation, and the amplification is increased. Consideration of the expression for ideal gain, $VA = 1.414GV_{T_r}$, shows that in the experimental case noted, placing $C_T = C_{N_r}$ results in double the gain possible when no tuning capacity is used. Although G is decreased by a factor of 2.8 when one-tenth this value of C_T is inserted, V_{T_r} is increased by a factor of 11 so that the over-all improvement in voltage gain is about 400 per cent.

The opportunity to control large tuned circuit voltages which is afforded by the inclusion of the tuning capacitor assumes even more importance when it is noted that the magnitude of the carrier voltage appearing across the nonlinear capacitor is seriously limited by various factors. As the voltage across the element becomes large the element loses percentage control of the carrier current. Harmonic generation in the circuit is greatly accentuated by increased carrier potential swings, and circuit stability suffers due to larger dielectric capacitor losses. Finally, erratic puncture, edge sparking, and fracture are found to occur at surprisingly low values of carrier potential with materials presently available.

In experimental work completed thus far the greatest circuit gain observed has been four. This gain was obtained under circumstances which approached the upper limit on operating conditions for the dielectric material. Maintaining the carrier voltage across the nonlinear condenser at a more reasonable value has given gains of at least three for the several amplifiers which have been constructed. The frequency response of a representative amplifier with a high impendance load is shown in Figure 21.

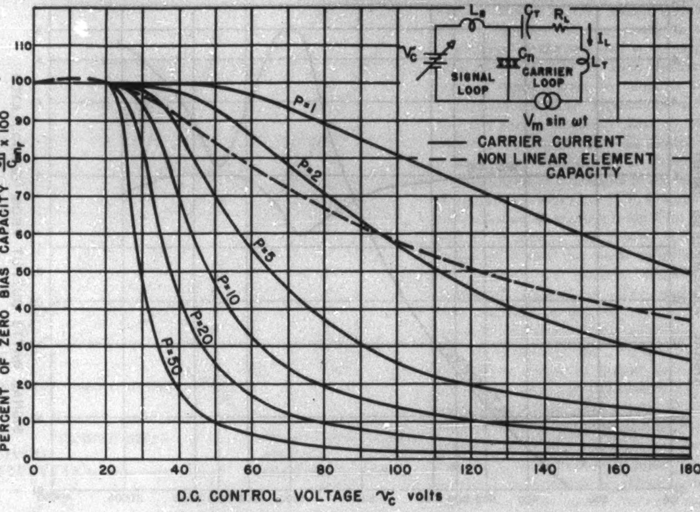
Figure 22 shows the tuned circuit modified for use as a general stage in a cascade arrangement of dielectric voltage amplifiers. In this circuit a portion of the direct plus signal voltage appearing at the carrier filter output is picked off by the potentiometer R_b. C_b acts to by-pass the signal component in this voltage and the direct component appears across C_N, thus eliminating the need for a bias battery. The steady-state response of an amplifier consisting of three such stages is given in Figure 23.

At present the investigation of the voltage amplifier is being continued along several lines. Consideration is being given to the real and reactive power output capabilities of the circuit, to improvement in amplifier frequency response, to the use of positive and negative feedback, and to a more exact analysis of nonlinear element resonant circuit characteristics.

List of Symbols

$V_m \sin \omega t$ = harmonic potential applied to load or carrier circuit
$V = 0.707V_m$

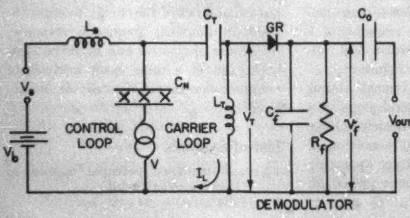
Figure 18. Detuning characteristics compared with the capacitance change of the nonlinear element, where $P=(\omega L_T/R_L)/(C_{Nr}/C_T+1)$

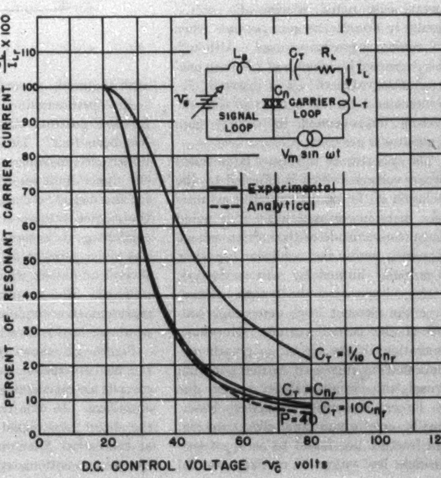


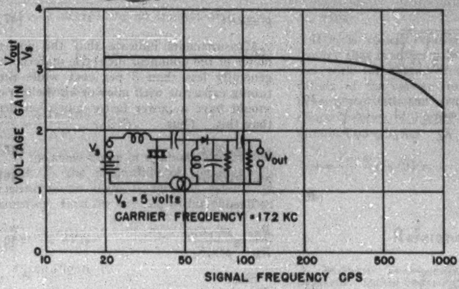
V_b = d-c bias potential
V_s = rms value of periodic signal voltage
v_s = instantaneous value of signal voltage
v_N = instantaneous voltage across nonlinear capacitor
v_c = instantaneous value of control voltage
v_f = instantaneous voltage across carrier filter
v_T = rms voltage across linear tuning element
V_{out} = rms output voltage
E = instantaneous value of potential gradient
i_L = instantaneous value of load or carrier circuit current
i_c = instantaneous value of control circuit current
I_{m_n} = crest value of *n*th harmonic component of current
I_{d_n} = crest value of dimensionless *n*th harmonic component of current
I_L = rms value of load or carrier current
q = instantaneous value of charge on nonlinear capacitor
Q_0 = average value of charge on nonlinear capacitor
Q_{m_n} = crest value of *n*th harmonic component of charge
q_L = instantaneous value of load circuit circulating charge
q_c = instantaneous value of control circuit circulating charge
D = electric displacement = charge per unit area for parallel plate capacitor
R_L = load circuit resistance
R_B = control circuit resistance
R_b = bias circuit resistance
R_N = equivalent resistance to represent losses in nonlinear element
R_{CT} = equivalent resistance to represent losses in tuning capacitor
R_{LT} = equivalent resistance to represent losses in tuning coil
R_0 = input resistance
L_B = blocking inductance
L_T = linear tuning inductance
L_f, C_f, R_f = carrier filter components
C = linear capacitor in load circuit
C_N = differential capacitance of nonlinear capacitor
C_T = linear tuning capacitance
C_b = bias circuit capacitor
C_e = equivalent tuned circuit capacitance
C_0 = blocking capacitor
pf_N = power factor of the nonlinear element
pfc_T = power factor of the linear tuning capacitance
S = surface area of nonlinear capacitor

Figure 19 (below). Dielectric voltage amplifier

Figure 20 (right). Detuning characteristics of the nonlinear dielectric element resonant circuit







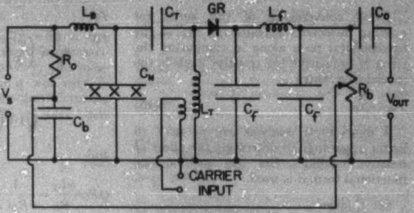


Figure 21 (left). Frequency response curve for a representative dielectric voltage amplifier

Figure 22 (above). General stage of a multistage dielectric voltage amplifier

A, a = constants determined from the relationship between D and E for a given dielectric material
l = distance between plates of nonlinear capacitor
φ = a phase angle
P = a factor which determines the shape of the detuning characteristics
QF = quality factor of the tuned carrier circuit
QF_L = quality factor of the linear tuning inductance
GR = germanium diode rectifier
Z_s = carrier source impedance
G = slope of detuning characteristic
VA = voltage amplification
Subscript r denotes circuit conditions at resonance

Appendix I. Derivation of General Characteristics

An analytical determination of the characteristics of a dielectric amplifier circuit must presume a functional form relating dielectric displacement and applied potential gradient. While this functional form can be made accurate to any desired degree, a proper balance must be struck between accuracy of representation and operational ease in solving the equations which result from this choice. The wide latitude of variation of this balance in modern analysis indicates that no small number of factors are involved; however, it may be said that complexity of analysis is an increasing function of accuracy.

Currently available dielectrics have been found to possess capacitances which vary from sample to sample and temperature to temperature in such fashion that any one of several functions may be used to advantage to represent the mean polarization curve. Suitable functions are

$$E = a_1D + a_3D^3 \tag{13}$$

$$D = A\tanh(aE) \tag{14}$$

$$D = A\sinh^{-1}(aE)^{3} \tag{15}$$

Equation 14 has been chosen in order to permit the following derivation of the no-load general characteristics of the dielectric amplifier.

Consider again the circuit of Figure 8. Equations of this circuit are

$$V_m \sin \omega t = q_L/C + R_L\dot{q}_L + v_N \tag{16}$$

$$v_c = -L_B\ddot{q}_c - R_B\dot{q}_c + v_N \tag{17}$$

$$q = q_L - q_c = SA\tanh av_N/l \tag{18}$$

In all cases which appear destined to find practical application

$$q_L >> q_c \tag{19}$$

because ωL_B is made very large compared to the greatest value of reactance of the nonlinear capacitor. Thus

$$q \cong q_L;\ q_c \cong 0 \tag{20}$$

No-load conditions require that

$$R_L = 0;\ \frac{1}{\omega C} \rightarrow \infty \tag{21}$$

which dictates that

$$R_L\dot{q}_L = 0;\ q_L/C = -v_c \tag{22}$$

Equations 16 and 17 become

$$V_m \sin \omega t + v_c = v_N \tag{23}$$

whence equation 18 can be written

$$q = SA\tanh\frac{a}{l}(V_m \sin \omega t + v_c) \tag{24}$$

This function can also be written as a Fourier series because of its periodicity.

$$q = Q_0 + Q_{m_1}\sin \omega t + Q_{m_2}\cos 2\omega t + Q_{m_3}\sin 3\omega t + \ldots \tag{25}$$

Here Q_{2K} and Q_{2KH} are functions of both V_m and v_c. It should be noted that the general dimensionless currents I_{d_n} are the quotients of Q_{m_n} and SA.

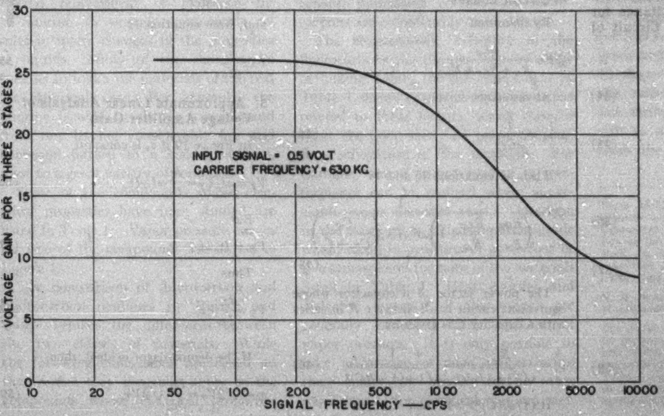


Figure 23. Frequency response curve for a 3-stage dielectric voltage amplifier

Rapid convergence of this series for existing dielectrics permits the use of the fundamental term alone in computing the effect of capacitive and resistive loads in the special case where

$$V_m << v_c \tag{26}$$

A quasi-linear analysis can be used to better advantage to determine the effect of resistive loading. For this method the functional form 3 is used.

For C large

$$v_c + V_m \sin \omega t = v_N + R_L \dot{q} \tag{27}$$

But since

$$\dot{q} = \frac{d}{dt}\left[SA \sinh^{-1} \frac{a v_N}{l} \right] \tag{28}$$

$$v_c + V_m \sin \omega t = v_N + \frac{R_L S A a}{\sqrt{a^2 v_N^2 + l^2}} \frac{dv_N}{dt} \tag{29}$$

In view of equation 26

$$v_c + V_m \sin \omega t = v_N + \frac{R_L S A a}{\sqrt{a^2 v_c^2 + l^2}} \frac{dv_N}{dt} \tag{30}$$

which is linear in v_N and has as its steady-state solution

$$v_N = v_c + \frac{V_m}{\sqrt{1 + \frac{\omega^2 R_L^2 S^2 A^2 a^2}{a^2 v_c^2 + l^2}}} \sin(\omega t + \varphi) \tag{31}$$

where

$$\varphi = \tan^{-1}\left[-\frac{R_L \omega S A a}{\sqrt{a^2 v_c^2 + l^2}} \right] \tag{32}$$

From this expression for the instantaneous voltage of the capacitor, the current i_L is readily determined, since

$$i_L = \frac{V_m \sin \omega t - (v_N - v_c)}{R_L} \tag{33}$$

Appendix II

1. Variation in Carrier Current with Nonlinear Element Capacitance for Series Resonant Carrier Circuit of Figure 17

Let

$$R_L \equiv R_N + R_{C_T} + R_{L_T} \tag{34}$$

$$C_e \equiv \frac{C_N C_T}{C_N + C_T} \tag{35}$$

For the series circuit

$$V = I_L \sqrt{R_L^2 + \left(\omega L_T - \frac{1}{\omega C_e} \right)^2} \tag{36}$$

$$I_L = \frac{V}{R_L} \frac{1}{\sqrt{1 + \left(\frac{\omega L_T}{R_L} - \frac{1}{\omega R_L C_e} \right)^2}} \tag{37}$$

At resonance

$$I_{L_r} = \frac{V}{R_L} \tag{38}$$

Hence

$$\frac{I_L}{I_{L_r}} = \frac{1}{\sqrt{1 + \left(\frac{\omega L_T}{R_L} - \frac{1}{\omega R_L C_e} \right)^2}} = \frac{1}{\sqrt{1 + \left(\frac{\omega L_T}{R_L} - \frac{1}{\omega R_L C_{e_r}} \frac{C_{e_r}}{C_e} \right)^2}} \tag{39}$$

But

$$QF \equiv \frac{\omega L_T}{R_L} = \frac{1}{\omega R_L C_{e_r}} \tag{40}$$

Thus

$$\frac{I_L}{I_{L_r}} = \frac{1}{\sqrt{1 + QF^2 \left(\frac{C_{e_r}}{C_e} - 1 \right)^2}} \tag{41}$$

Now

$$\frac{C_{e_r}}{C_e} = \left(\frac{C_{N_r} C_T}{C_{N_r} + C_T} \right) \Big/ \left(\frac{C_N C_T}{C_N + C_T} \right) \tag{42}$$

$$\left(\frac{C_{e_r}}{C_e} - 1 \right) = \frac{C_{N_r} C_T (C_N + C_T) - C_N C_T (C_{N_r} + C_T)}{C_N C_T (C_{N_r} + C_T)} = \frac{\frac{C_{N_r}}{C_N} - 1}{\frac{C_{N_r}}{C_T} + 1} \tag{43}$$

so that

$$\frac{I_L}{I_{L_r}} = \frac{1}{\sqrt{1 + \left[\frac{QF}{\frac{C_{N_r}}{C_T} + 1} \right]^2 \left(\frac{1}{\frac{C_N}{C_{N_r}}} - 1 \right)^2}} = \frac{1}{\sqrt{1 + P^2 \left(\frac{1}{\frac{C_N}{C_{N_r}}} - 1 \right)^2}} \tag{11}$$

2. Expression for P when Nonlinear Element Losses Dominate Carrier Circuit Losses

By definition

$$QF = \frac{\omega L_T}{R_N + R_{C_T} + R_{L_T}}$$

At resonance

$$\omega L_T = \frac{1}{\omega C_{e_r}} \tag{44}$$

Then by equations 35 and 44

$$QF = \frac{1}{\left(\frac{R_{C_T}}{R_N} + \frac{R_{L_T}}{R_N} + 1 \right) \omega R_N C_{N_r} \left(\frac{C_T}{C_{N_r} + C_r} \right)} \tag{45}$$

The power factor of a capacitor whose equivalent circuit is a resistance R in series with a capacity C is given by

$$pf = \frac{R}{\sqrt{R^2 + 1/\omega^2 c^2}} = \frac{1}{\sqrt{1 + (1/\omega RC)^2}} \tag{46}$$

If $(1/\omega RC)^2 >> 1$, then

$$pf \cong \omega RC \tag{47}$$

Measurements indicate that the power factor of the nonlinear dielectric material is generally less than 3 per cent while the tuning capacitor with mica or air dielectric should have a power factor much smaller than this. Thus

$$pf_{N_r} \cong \omega R_N C_{N_r} \tag{48}$$

$$pf_{C_T} \cong \omega R_{C_T} C_T \tag{49}$$

Then

$$\frac{R_{C_T}}{R_N} \cong \frac{pf_{C_T} C_{N_r}}{pf_{N_r} C_T} \tag{50}$$

Since

$$QF_L \equiv \frac{\omega L_T}{R_{L_T}} \tag{51}$$

$$\frac{R_{L_T}}{R_N} \cong \frac{\omega L_T \omega C_{N_r}}{QF_L pf_{N_r}} \frac{C_{e_r}}{C_{e_r}} = \frac{1}{QF_L pf_{N_r}} \times \frac{C_{N_r}(C_{N_r} + C_T)}{C_{N_r} C_T} = \frac{\left(\frac{C_{N_r}}{C_T} + 1 \right)}{QF_L pf_{N_r}} \tag{52}$$

Substituting equations 48, 50, and 52 into equation 45

$$QF \cong \frac{1}{\left[\frac{pf_{C_T} C_{N_r}}{pf_{N_r} C_T} + \frac{\left(\frac{C_{N_r}}{C_T} + 1 \right)}{pf_{N_r} QF_L} + 1 \right] \times pf_{N_r} \left[1 \Big/ \left(\frac{C_{N_r}}{C_T} + 1 \right) \right]} \tag{53}$$

If components are chosen such that

$$\frac{R_{C_T} + R_{L_T}}{R_N} = \left[\frac{pf_{C_T} C_{N_r}}{pf_{N_r} C_T} + \frac{\left(\frac{C_{N_r}}{C_T} + 1 \right)}{pf_{N_r} QF_L} \right] << 1 \tag{12}$$

then

$$QF \cong \frac{\left(\frac{C_{N_r}}{C_T} + 1 \right)}{pf_{N_r}} \tag{54}$$

and, from equation 11

$$P \cong \frac{1}{pf_{N_r}} \tag{55}$$

3. Approximate Linear Analysis of Voltage Amplifier Gain

In Figure 19 if v_c is constant

$$V_T = \omega I_L L_T = \frac{I_L}{I_{L_r}} \omega L_T I_{L_r} \tag{56}$$

But

$$V_{T_r} = I_{L_r} \omega L_T \tag{57}$$

Thus

$$V_T = \frac{I_L}{I_{L_r}} V_{T_r} \tag{58}$$

If the demodulator is ideal, then

$$v_f = \sqrt{2} V_T = \frac{I_L}{I_{L_r}} \sqrt{2} V_{T_r} \tag{59}$$

If v_e is changed by an amount Δv_e

$$\Delta v_f = \Delta\left(\frac{I_L}{I_{L_r}}\right)\sqrt{2}V_{T_r} \qquad (60)$$

Dividing by Δv_e

$$\frac{\Delta v_f}{\Delta v_e} = \frac{\Delta(I_L/I_{L_r})}{\Delta V_e}\sqrt{2}V_{T_r} \qquad (61)$$

If for some range in v_e about a fixed bias point V_b the magnitude of the detuning characteristic slope G remains essentially constant, then for any Δv_e within this range

$$\frac{\Delta v_f}{\Delta v_e} = G\sqrt{2}V_{T_r} \qquad (62)$$

From Figure 19

$$v_e = v_s + V_b \qquad (63)$$

$$\Delta v_e = \Delta v_s \qquad (64)$$

If v_s is varied sinusoidally with a frequency that is sufficiently small in comparison with the carrier frequency, the magnitude of the carrier current will vary in accordance with the steady-state detuning characteristics of Figure 18. Under these circumstances

$$VA = \frac{V_{out}}{V_s} = G\sqrt{2}V_{T_r} \qquad (65)$$

References

1. Electric Translation Circuit, **Gustave Guanella.** United States Patent *2,191,315*, February 1940.

2. Circuit Arrangement for Modulating an Electric Signal, **Gerard Hepp.** United States Patent *2,470,893*, May 1949.

3. Ferroelectricity, Domain Structure, and Phase Transitions of Barium Titanate, **A. von Hippie.** *Review of Modern Physics* (New York, N. Y.), volume 22, number 3, July 1950, pages 221–37.

4. Analytical Determination of Characteristics of Magnetic Amplifiers with Feedback, **D. W. Ver Planck, L. A. Finzi, D. C. Beaumariage.** *AIEE Transactions*, volume 68, part 1, 1949, pages 565–70.

5. An Analysis of Transients and Feedback in Magnetic Amplifiers, **W. C. Johnson, W. F. Latson.** *AIEE Transactions*, volume 69, part I, 1950, pages 604–12.

6. General Characteristics of Magnetic Amplifiers, **L. A. Finzi, D. C. Beaumariage.** *AIEE Transactions*, volume 69, part I, 1950, pages 919–27.

No Discussion

Some Fluorinated Liquid Dielectrics

N. M. BASHARA
NONMEMBER AIEE

Synopsis: The properties of some fluorinated liquids are reviewed. They have low dielectric constants and low power factors over wide frequency and temperature ranges. They also have high dielectric strengths. A range of boiling points is available to make possible high vapor densities at relatively low temperatures. Utilization of this in electrical applications is feasible with these compounds since they meet requirements of nonflammability, stability, and compatibility with metals and insulation which are not satisfied entirely by other materials. Low water absorption and high thermal stability are additional reasons for interest in these materials.

THE replacement of hydrogen by fluorine in organic compounds results in many changes in the properties of liquids. Many of the changes are such as to make the materials of interest for electrical uses. For example, the fluorine is able to change a compound from one having considerable fire and explosion hazard to a stable material, inert to a great variety of conditions.

Some of the compounds whose electrical properties have been studied are listed in Table I. Vapor pressure curves for two of the compounds are shown in Figure 1.

The comparison of fluorocarbon and hydrocarbon pentanes by Burger and Cady[1] typifies the differences between the two classes of materials. While the following conclusions are based on that report they generally apply to the compounds whose electrical properties have been studied here. It should be emphasized that the similarity in properties to fluorocarbons is being stated as a generalization only for the particular liquids studied in this report. Attention is called to this point by referring to these specific materials as fluorinated liquids. Comparing similar compounds the fluorocarbons have a greater density. Their absolute viscosity is greater but the kinematic viscosity is less. The fluorocarbons boil at lower temperatures. The coefficient of expansion of the fluorocarbons is greater, as is the change of density and viscosity. The over-all general impression is that the fluorocarbons are considerably more volatile.

The characteristic volatility of the fluorocarbons plus the generally low range of boiling points for the compounds of Table I direct emphasis to applications related to these factors, which complement the good electrical properties and fireproof nature of the materials. For example, many improvements in performance can be realized where appreciable vapor densities exist. Operation in the range up to 200 degrees centigrade would result in pressures of up to four or five atmospheres for some of the materials listed in Table I. Heat transfer and dielectric strength of the vapors are generally benefited with increase of vapor pressure. It is only possible to take advantage of such properties in these compounds because certain other requirements for electrical uses are met. In contrast, low boiling hydrocarbons and chlorinated compounds do not find such uses, failing to meet one or more of the requirements. Hydrocarbons are unsatisfactory due to the aggravated fire hazard of the low boiling compounds. Low boiling chlorinated materials, where available, are objectionable due to their deteriorating effect on many types of insulation and metals and because of their toxicity. The fluorinated liquids have none of these objections. They are also much more stable to other conditions.

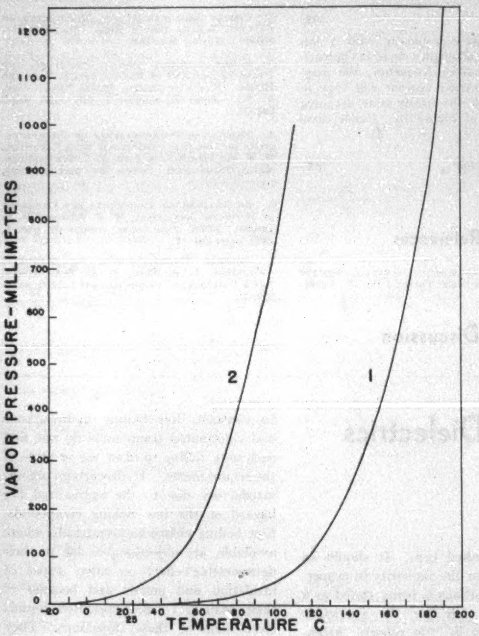
Obviously, for the same reasons that new design variations become possible with these materials there must be alterations in design to accomodate them. The liquids discussed here require a closed system to prevent evaporation. More room for expansion is required, this becoming even more of a problem at elevated temperatures. The coefficient of volume expansion increases by 75 per cent for $(C_4F_9)_3N$ at 150 degrees centigrade and by 25 per cent for c-$C_8F_{16}O$ at 75 degrees centigrade as compared to the values at room temperature. The low surface tension makes it more difficult to make a tight seal. In certain cases the increased weight for a given

---

Paper 53-135, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted September 30, 1952; made available for printing December 1, 1952.

N. M. Bashara is with the Minnesota Mining and Manufacturing Company, St. Paul, Minn.

The writer wishes to express his thanks to R. Herr for the heat-transfer data of Figure 8 and to members of Central Research and the Fluorochemical Department for their help in obtaining the data of Table I and many other instances of advice and co-operation. The sample materials were provided by the Central Research Department.



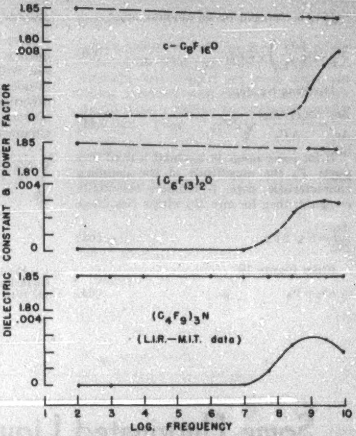


Figure 1 (left). Vapor pressure of 1-$(C_4F_9)_3N$ and 2-c-$C_8F_{16}O$

Figure 2 (above). Dielectric constant and power factor as a function of frequency for the three types of compounds

Measurements on $(C_4F_9)_3N$ courtesy Laboratory for Insulation Research, Massachusetts Institute of Technology

volume might be objectionable. The large temperature coefficients of density and viscosity are also factors to be considered.

Electrical Properties

Table II, which lists some electrical properties at 25 degrees centigrade, indicates that these liquids are unusual, with very low dielectric constants and refractive indices, low conductivities, and power factors which, it will be seen, remain small over a wide range of frequencies.

The average values of dielectric strength are somewhat higher on the American Society for Testing Materials test than those ordinarily listed for transformer oils. However, a comparative check of equally pure samples of these liquids and transformer oils has not been made.

Since there is a difference between the square of the refractive index and the low frequency dielectric constant, energy absorption is expected. Figure 2, which plots the dielectric properties to the microwave region, shows the presence of dissipation factor peaks at the upper end of the spectrum. The tests were made on samples which had undergone considerable purification. The relatively small change of dielectric constant (less than 1 per cent) to the upper microwave range anticipates a large change in the infrared which is confirmed experimentally.[2,3]

Some of the effects of temperature and moisture on the electrical behavior have been determined. There was no consistent change in the dielectric strength between 30 and 90 degrees centigrade for dry samples. Figure 3 shows the decrease in resistance with temperature for typical, dry samples of two of the liquids with different boiling points. There is a decrease of dielectric constant of approximately 0.0016 per degree for c-$C_8F_{16}O$ between 30 and 90 degrees centigrade and about 0.0012 per degree for $(C_6F_{13})_2O$ between 30 and 150 degrees centigrade.

Table III summarizes the effects of water. There appears to be a larger effect on the electrical properties of these materials than on commonly used trans-

Table I. General Properties of Some Fluorinated Liquids*

Material	Pour Point, Deg. C., ASTM D-97[10]	Boiling Point, Deg. C.	Density, 25 Deg. C., Grams per Centimeter	Viscosity, 25 Deg. C., Centistokes	Surface Tension, 25 Deg. C., Dynes per Centimeter	Volume Expansion, per Deg. C., at 25 Deg. C.	Specific Heat, Calories/Grams Deg. C.	Heat of Vaporization Calories/Grams
$(C_2F_5)_3N$		69	1.73	0.501				
$(C_3F_7)_3N$		129	1.82	0.806				
$(C_4F_9)_3N$	− 50	178	1.87	2.96	16.1	1.2×10^{-3}	0.27	16.5
$(C_4F_9)_2O$		101	1.71	0.612	13.0			
$(C_6F_{13})_2O$	− 90	172	1.81	2.11	16.3			
c-$C_6F_{12}O$		56	1.69	0.475				
c-$C_8F_{16}O$	< −100	103	1.77	0.809	15.2	1.6×10^{-3}	0.26	20.9

* Commercial grade materials could show some deviation from these results. These data from reference 11.

Table II. Electrical Properties of Some Fluorinated Liquids at 25 Degrees Centigrade*

Material	Dielectric Constant, 100c	Dielectric Constant, 10^{10}c	Dissipation Factor, 100c	Refractive Index	Resistivity, Ohm-Centimeters, (2 Kv per Centimeter)	Dielectric Strength, Kv, ASTM D-877[19]
$(C_2F_5)_3N$	1.89		<0.0005	1.261	10^{16}–10^{17}	39
$(C_3F_7)_3N$	1.85		<0.0005	1.279	10^{15}–10^{17}	44
$(C_4F_9)_3N$	1.86	1.85	<0.0005	1.290	10^{14}–10^{16}	45
$(C_4F_9)_2O$	1.77		<0.0005	1.261	10^{15}–10^{17}	43
$(C_6F_{13})_2O$	1.85	1.84	<0.0005	1.278	10^{16}–10^{17}	45
c-$C_6F_{12}O$	1.85		<0.0005	1.263	10^{16}–10^{17}	41
c-$C_8F_{16}O$	1.85		<0.0005	1.278	10^{15}–10^{17}	43

* Commercial grade materials could show some deviation from these results which, except for the amines, were made on relatively pure samples. The purity of the amines was uncertain.

Table III. Effect of Water on Dielectric Strength and Resistivity*

Test Sample		Resistivity (Ohm-Centimeters) Dry	Resistivity (Ohm-Centimeters) 90% Relative Humidity	Dielectric Strength, Kv Dry	Dielectric Strength, Kv 90% Relative Humidity	PPM Water, 90% Relative Humidity
c-$C_8F_{16}O$	a†	$<10^{16}$	4×10^{16} (80 days)	40	17 (80 days)	
	b	$<10^{16}$	5×10^{13} (10 days)	40	17 (10 days)	
$(C_6F_{13})_2O$	a	$<10^{16}$	3×10^{11} (80 days)	40	17 (80 days)	~20‡
	b	$<10^{16}$	2×10^{12} (10 days)	40	15 (10 days)	~10
$(C_4F_9)_3N$	b	$<10^{16}$	4×10^{13} (10 days)	40	15 (10 days)	

* All tests at approximately 25 degrees centigrade. Dielectric strength test, *ASTM D-877.*[19] Resistance measurements at voltages 2 kv per centimeter.

† Two different samples of c-$C_8F_{16}O$ and $(C_6F_{13})_2O$ were tested. Results for samples (a) are at the end of 80 days, samples (b) at the end of 10 days.

‡ Absolute accuracy poor at these low concentrations since measurement was made by Karl Fischer method. The difference between the two samples is probably real. There was good reproducibility on a large number of tests.

former oils even though the maximum water saturation appears to be no greater than 25 parts per million for these liquids at 90-per-cent relative humidity. At this same humidity a hydrocarbon transformer oil would show 73 parts per million and an Askarel 108 parts per million. These results are from Clark.[4] His tests at room temperature and a water content corresponding to 90-per-cent relative humidity show that the dielectric strength of the Askarel dropped to about 40 per cent of the dry value. Mineral oil showed a decrease to about 80 per cent of the dry value. The fluorinated liquids tested in Table III decreased in dielectric strength in about the same ratio as the Askarels for the same test conditions. The resistance also shows a greater decrease for the fluorinated liquids despite their lower water absorption. The apparently greater change of resistance may be due to effects related to lower viscosity, such as ion mobility, and so forth. The solubility of water in perfluorodimethylcyclohexane was found to be 13.6 parts per million at 25 degrees centigrade.[5,6]

It might be well to point out that the comparison of the effect of water has neglected what is in practice a more serious problem for hydrocarbon oils, condensed water which can accumulate at the bottom of apparatus. No such problem would occur in the use of the fluorinated liquids since there would be no accumulation of condensed vapor. Comparative tests on the dielectric strength of a hydrocarbon oil provided an example of what can occur. The oil sample, while exposed to 90-per-cent relative humidity, underwent slight but sufficient temperature cycling to result in accumulation of condensed vapors. The test samples contained sufficient amounts of occluded water to lower the dielectric strength to about one-fourth of the dry value. The overall comparison should certainly take this into account since the total amount of water which can accumulate in a hydrocarbon oil under temperature cycling is considerably greater than that obtained when the sample is exposed to a given humidity at a constant temperature. On the other hand, for the fluorinated liquids, only the maximum saturation for any given humidity could occur. Water droplets did not form even when a layer of water was placed over the test material.

Because of the small amount of water absorbed by the fluorinated liquids, the apparatus used in the resistivity and dielectric strength tests were conditioned at the humidity under test to avoid loss of water to the equipment.

An inverted-V type curve is obtained for the change of resistance with temperature for a sample containing water. In this case the test apparatus was not maintained at a constant high humidity during this test. Diffusion is rapid in these liquids. The characteristic curve obtained may indicate the loss of the small amount of water present and a rise in resistance until increased ion mobility causes a decline to take place. A hydrocarbon oil shows the same behavior although the mechanism is probably different.[7] There the initial rise is explained as indicating a transition from water in suspension to a condition where increasing amounts go into solution.

With the range of boiling points available, considerable vapor densities can exist at relatively low temperatures, depending on the choice of material. For example, while the pressure of the vapor above a hydrocarbon transformer oil is very low at a temperature of 90 degrees centigrade, many of the fluorinated liquids studied would have an appreciable pressure at this temperature.

Figure 4 shows the differences in dielectric strength of the vapors of two of the liquids. Comparison is also made with a fluorocarbon gas and nitrogen. The vapors of carbon tetrachloride would show a dielectric strength similar to c-$C_8F_{16}O$ on Figure 4.

High vapor densities at relatively low

Table IV. Effect of c-$C_8F_{16}O$ and $(C_6F_{13})_2O$ on Metals and Insulation at 90 Degrees Centigrade*

		Iron	Aluminum	Copper	Silicone Rubber	Varnished Cambric	Teflon	Kraft Paper
Air present	Vapor Phase	NC†	NC	NC	NC	NC	NC	NC
(1 week and 11 weeks)	Liquid Phase	NC	NC	NC	NC	NC	NC	NC
Water present, no air	Vapor Phase	Sev. Cor.	Sl. Cor.	Sl. Cor.	NC	Varn. Ext.	NC	NC
(1 week and 11 weeks)	Liquid Phase	Sl. Cor.	Sl. Cor.	NC	NC	Varn. Ext.	NC	NC

* The test sample was partly immersed in the liquid. Two sets of tests were made for each liquid. There was approximately 50 per cent relative humidity air above the liquid in one series and a layer of water above the liquid in the other. Little or no change occurred after the first week.

† NC—no change.

temperatures can be obtained from the two fluorinated liquids of Figure 4, c-$C_8F_{16}O$ and $(C_4F_9)_3N$. The vapor phase above the liquid can therefore have an appreciable dielectric strength in the temperature range of 50 to 150 degrees centigrade. In contrast, the vapors of the commonly used transformer oils would have little or no dielectric strength at these temperatures because of low vapor densities. Low boiling hydrocarbons and chlorinated liquids which would allow high vapor concentrations are undesirable for the reasons given.

Stability

There are basic reasons related to molecular structure for expecting the fluorocarbons to be stable materials. Experimental verification of this has been obtained in a large number of experiments. A good treatment was recently published by Stuenenberg and Cady[8] which deals with the heat stability of fluorocarbons. Their results showed that the fluorocarbons tested were more stable than the corresponding hydrocarbons, decreasing in the order CF_4, C_2F_6, C_3F_8, and so forth. Figure 1 in their article gives a concise summary.

Tests have been done on many of the compounds listed in Table I by passing them through carbon-lined tubes. Appreciable decomposition does not occur until temperatures of 450 to 600 degrees centigrade are reached. At temperatures above 500 degrees centigrade some toxic products are found.

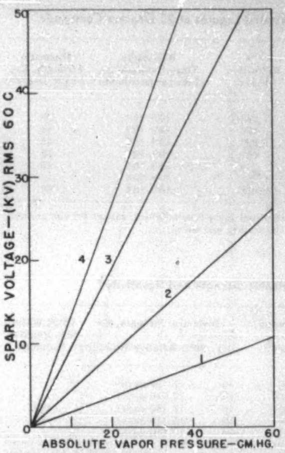


Figure 4. Dielectric strength of the vapors of:

1. Nitrogen
2. C_3F_8
3. c-$C_8F_{16}O$
4. $(C_4F_9)_3N$

0.25-inch diameter spheres, 0.25-inch gap

Since in certain types of operation the liquid is subject to arcing, it is necessary to know about the behavior of these materials under such conditions. Decomposition due to the high temperature of the arc is, of course, expected. The by-products formed are obviously related to the metals and insulation present which can react when free radicals are produced. Strips of commonly used materials were included in the arc tests. In arcing the types of fluorinated liquids studied here lower molecular weight saturated and unsaturated fluorocarbons were produced as well as CO, CO_2, COF_2, and various fluorides, probably including acid fluorides. These acid fluorides would be expected to have a deteriorating effect on metals and insulation. No by-products which are known to be toxic were identified. Although every attempt was made to trap all by-products, until extensive testing is done the absence of toxic components cannot be an absolute certainty.

The effects of corona on gases have been studied.[9] Breakdown into lower and some higher molecular weight products was observed. No toxic products were noted. Perfluoropropane was found to yield smaller amounts of corrosive products than sulphur hexafluoride, which is itself a very stable compound. Fluorocarbons would be expected to be more stable to corona than the Freons.

Differences in the amount of carbon produced from arcing various fluorinated liquids, a hydrocarbon transformer oil, and a chlorinated oil are indirectly shown in Figure 5. A relatively low current arc (0.050 ampere) was initiated by high voltage breakdown. The light transmission after arcing is a measure of the darkening due to formation of carbon. The Askarel carbonized most severely, followed by the hydrocarbon oil. The fluorinated liquids show the least carbonization. It should be pointed out that the decomposition products from these fluorinated liquids and the Askarels are more objectionable than those obtained from a hydrocarbon in one respect. Products could be obtained which would have a deteriorating effect on insulation and metals, such as hydrogen chloride and acid fluorides. Of course, the breakdown of a hydrocarbon can produce hydrogen, accompanied by the explosion hazard.

Table IV summarizes tests on metals and plastics sealed in glass ampoules at 90 degrees centigrade. The greatest effect was observed in the ampoule containing a sample of varnished cambric. Extraction and darkening of the varnish was observed. The small amount of water present in the liquid is evidenced by severe corrosion of the iron in the vapor phase and only slight corrosion of the part immersed in the liquid.

Pure samples of c-$C_8F_{16}O$ have been tested in a stainless steel autoclave at 300 degrees centigrade to determine the thermal stability. There was no detect-

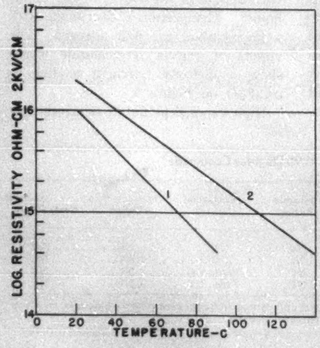
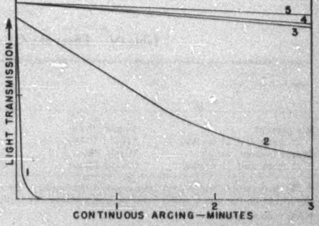


Figure 3 (left). Change of resistance with temperature of two of the liquids (dry samples)

1. c-$C_8F_{16}O$
2. $(C_6F_{13})_2O$

Figure 5 (right). Carbonization of various liquids under arcing. Approximately 0.050 ampere alternating current

1. Askarel transformer oil
2. Hydrocarbon transformer oil
3. $(C_6F_{13})_2O$
4. c-$C_8F_{16}O$
5. c-$C_6F_{12}O$



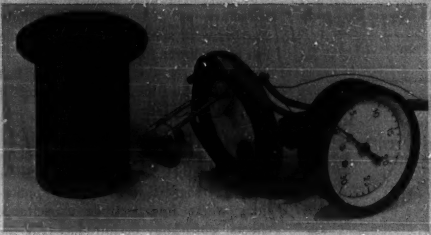


Figure 6 (above). Apparatus used in heat-transfer testing

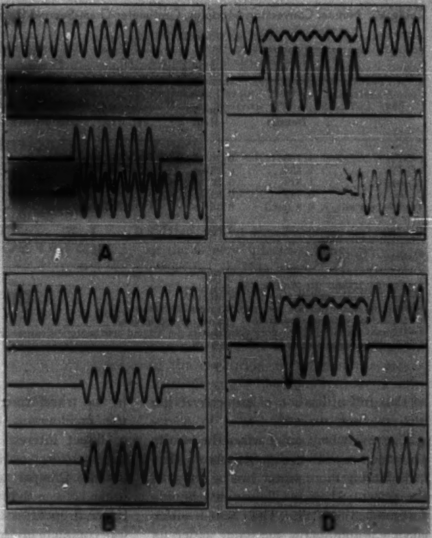


Figure 8 (right). Circuit-breaker tests. The curves are: top—115-volt line; second—circuit-breaker current; bottom—voltage across circuit-breaker contacts, 2,600 volts fully opened

A. c-$C_8F_{16}O$—275 amperes
B. c-$C_8F_{16}O$—900 amperes
C. c-$C_8F_{16}O$—3,450 amperes
D. Hydrocarbon transformer oil—3,450 amperes

able breakdown of the material. Not more than 1 part per million of fluoride was present, which was the lower limit of the analytical method.

All of the materials reviewed here are nonflammable, and this is probably true for all completely fluorinated liquids. Carbon tetrafluoride has been proved to be an effective inhibitor of explosions over a considerable range in tests on anesthetics.[10] Isolated tests involving arcing mixtures of 6 and 25 per cent oxygen in C_2F_6 produced no explosions at room temperature. No explosions have resulted in passing vapors of these and other fluorocarbons through hot tubes up to temperatures of 1,000 degrees centigrade even in the presence of steam. These materials show little or no solubility with hydrocarbon and chlorinated solvents and oils.[11]

Applications

The properties of these materials suggest their use in new and unusual ways. For the first time there are available stable, inert liquids which cover a range of boiling points up to 178 degrees centigrade and which have no fire hazard in either the vapor or liquid phase. Neither hydrocarbon nor chlorinated liquids have been satisfactory on all points, the hydrocarbons because of their flammability and the chlorinated compounds because of their solvent action. For such applications certain limitations of low boiling materials must be accepted, as for instance the need for closed-system operation. However, advantage can be taken of a coexistent factor such as high vapor density.

HEAT TRANSFER

The best-known electrical application to date is the Westinghouse patent for cooling power transformers by utilizing the heat of vaporization.[12] Other operating requirements of high dielectric strength of the vapors, thermal stability, and compatability with metals and insulation are met.

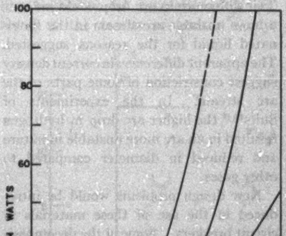


Figure 7. Efficiency of heat transfer

1. Silicone oil, high viscosity
2. Transformer oil
3. $(C_4F_9)_3N$
4. c-$C_8F_{16}O$
5. c-$C_6F_{12}O$

Next to heat of vaporization, convection is generally the most effective technique of heat removal. The compounds studied in this report have superior autoconvection properties. Table V lists the natural convection modulus[13] for various materials. While the heat transfer is not directly proportional to the differences in modulus, the compounds studied here would all be expected to be better than transformer oil when autoconvection is utilized. The improved heat transfer by autoconvection arises mostly out of differences in density, viscosity, and coefficient of expansion. While it might reasonably be argued that an improvement in heat transfer by autoconvection can also be realized by using low boiling hydrocarbons and chlorinated compounds, it is impractical to do so for the reasons outlined. An analysis by Robinson[14] relates both theoretically and experimentally those properties of cooling media, including liquids, to heat removal from practical electronic constructions.

Figure 6 shows equipment used in heat-transfer tests. The test liquid surrounds a small radio-frequency coil. Obviously, the apparatus is not of the best design to take full advantage of streamline flow

Table V. Natural Convection Modulus, Centimeter-Gram-Second Units, 25 Degrees Centigrade

$$a = \frac{p^2 g c_p b^*}{kn}$$

	Density	Specific Heat	Coefficient Vol. Exp.	Thermal Conductivity	Kinematic Viscosity (n/p)	Nat. Con. Modulus
c-$C_8F_{16}O$	1.760	0.26	1.6×10^{-3}	27.6×10^{-5}	0.82×10^{-2}	317,000
c-C_8F_{16}	1.835	0.243	1.36	30	0.975	202,000
$(C_4F_9)_3N$	1.872	0.27	1.2	25 (est.)	2.74	87,000
$C_{14}H_{30}$	0.732	0.523	0.893	33	3.0	34,000
Transformer oil	0.8 (est.)	0.45 (est.)	1.0 (est.)	27	21	7,000 (approx.)
DC 200	0.971	0.35	1.6	37	200	850

* where

a = natural convection modulus(1 per centimeter^3, degrees centigrade)
p = density (grams per centimeter^3)
g = gravity constant (centimeters per second^2)
c_p = specific heat (calories per gram, degrees centigrade)
b = coefficient volume expansion (1 per degree centigrade)
k = thermal conductivity (calory-centimeters per second-centimeters^2-degrees centigrade)
n = viscority (dyne-seconds per centimeter^2)

so that full utilization of autoconvection has not been realized.

Figure 7, which summarizes the results obtained on various liquids, shows that considerably more power can be put into the system when the fluorinated liquids surround the coil. The heat-transfer efficiency follows in the same order as the autoconvection factor. For the system used an increase in efficiency over transformer oil of from 50 to 300 per cent or more is possible.

CIRCUIT BREAKERS

Many of the characteristics of these liquids suggest their application to circuit breakers. Foremost is the fact that they are nonflammable. Also, the results of Figure 5 indicate less carbon formation with arcing than with hydrocarbon oils and Askarels. In addition, certain basic properties of fluorine in general and the volatility characteristic of the fluorinated liquids studied in this article favor their use. Thermal ionization for a given arc temperature would be less due to the higher ionization potential of fluorine. The greater recombination coefficient of fluorine would enter importantly into the neutralization of ions formed in the arc. Perhaps even more important, the volatile nature of many of the compounds would encourage the formation of eddies. This has been shown to be very important to cooling the arc stream.[15]

c-$C_8F_{16}O$ and a hydrocarbon transformer oil were compared in tests in a plain-type circuit breaker. The oval-shaped circuit-breaker tank required 6 gallons of liquid for filling. The tests were made on a 2,600-volt a-c system. Currents of 275, 900, and 3,450 amperes were interrupted in the fluorocarbon. The highest current was also interrupted in the transformer oil. Figure 8 shows the oscillograms obtained. The point of direct interest is that apparently a larger arc drop occurred in the fluorinated liquid. Comparing the 3,450-ampere test it can be seen that there is a considerably larger drop across the contacts just before complete interruption in the fluorinated liquid. Compare (C) and (D) on Figure 8. The bottom curves show the voltage drop across the contacts. Of particular interest is the voltage drop just before extinction. Of course, after extinction the voltage across the contacts is full line voltage. The arc drop is considerably greater in (C) than in (D). This difference in arc drop could indicate a more unstable arc stream in the fluorinated liquid for the reasons suggested. The apparent differences in current density suggest constriction of some parts of the arc stream. In the experiments of Suits[16,17] the higher arc drop in hydrogen resulted in an arc more unstable in nature and reduced in diameter compared to other gases.

New design problems would be introduced in the use of these materials in circuit breakers. Some of the decomposition products of arcing would attack insulating materials and metals. The amount of such products would depend on the number of arcs and the energy contained in them. Absorbents would be necessary for some conditions. Also, the density of these liquids is less favorable to the settling of carbon.

IMPREGNANTS

There is an acknowledged need for impregnants which are stable at temperatures above 150 degrees centigrade. As pointed out in the section entitled "Stability," there is no chemical breakdown of pure c-$C_8F_{16}O$ in a stainless-steel autoclave at 300 degrees centigrade. Further work needs to be done to establish upper limits for these materials.

Attention is also called to the fact that the surface tensions of these liquids are among the lowest known.

LOW CAPACITANCE

The low dielectric constants of these materials suggests their use in applications where interelectrode capacitance effects, which may be introduced by the liquid, need to be minimized.

DIELECTRIC STRENGTH OF VAPORS

The use of these liquids would give increased protection at relatively low ambient temperatures to the parts of equipment above the liquid. The high dielectric strength of the vapors would make it possible to operate at higher voltage gradients than are allowable with conventional transformer oils.

References

1. PHYSICAL PROPERTIES OF PERFLUOROPENTANES, **L. L. Burger, G. H. Cady.** *Journal*, American Chemical Society (Washington, D. C.), volume 73, September 1951, pages 4243–46.

2. THE INFRA-RED SPECTRA OF FLUORINATED HYDROCARBONS. I. **P. Tarkington, H. W. Thompson.** *Transactions*, Faraday Society (London, England), volume 41, 1945, pages 236–46.

3. THE INFRA-RED ABSORPTION SPECTRUM OF HEXAFLUOROETHANE GAS, **J. R. Nielsen, C. M. Richards.** *Journal of Chemical Physics* (New York, N. Y.), volume 16, January 1948, pages 67–73.

4. WATER SOLUTION IN HIGH-VOLTAGE DIELECTRIC LIQUIDS, **F. M. Clark.** *AIEE Transactions (Electrical Engineering)*, volume 59, August 1940, pages 433–41.

5. *Document MDDC 1455*, **J. L. Gabbard, E. S. Amis, A. V. Faloon, J. C. Pierce.** Atomic Energy Commission, Washington, D. C.

6. *Document MDDC 1454*, **J. L. Gabbard, E. S. Amis, J. H. Jamieson, T. J. McCormick.** Atomic Energy Commission, Washington, D. C.

7. THE CONDUCTIVITY OF HYDROCARBON TRANSFORMER OIL CONTAINING WATER AND SOLID CONDUCTING PARTICLES, **A. W. Stannett.** *British Journal of Applied Physics* (London, England), volume 2, April 1951, pages 110–14.

8. PYROLYSIS OF FLUOROCARBONS, **R. K. Steunenberg, G. H. Cady.** *Journal*, American Chemical Society (Washington, D. C.), volume 74, August 20, 1952, pages 4165–68.

9. SOME FLUOROCHEMICALS FOR ELECTRICAL APPLICATIONS, **N. M. Bashara.** *Proceedings*, Symposium on Progress in Quality Electronic Components (Washington, D. C.), May 1952, pages 30–35.

10. PRELIMINARY INVESTIGATIONS OF CARBON TETRAFLUORIDE AS AN INERT DILUENT GAS TO PREVENT EXPLOSIONS OF MIXTURES OF CYCLOPROPANE AND OXYGEN, **C. S. Jones, A. Faulconer, Jr., E. J. Baldes.** *Anesthesiology* (Philadelphia, Pa.), volume 11, September 1950, pages 562–66.

11. ELECTRICAL PROPERTIES OF INERT LIQUIDS. *Fluorochemical Brochures*, Minnesota Mining and Manufacturing Company (St. Paul, Minn.), April 17, 1952.

12. COOLING AND INSULATING ELECTRICAL APPARATUS, **C. F. Hill.** United States Patent *2,561,738*, July 24, 1951.

13. PHYSICS IN CHEMICAL INDUSTRY (book), **R. C. L. Bosworth.** Macmillan Company, New York, N. Y., 1950, pages 595–98.

14. Heat Transfer from Electronic Components, W. D. Robinson. *Proceedings*, Symposium on Progress in Quality Electronic Components (Washington, D. C.), May 1952, pages 125–32.

15. Displacement and Diffusion in Fluid-Flow Arc Extinction, Joseph Slepian. *AIEE Transactions* (*Electrical Engineering*), volume 60, April 1941, pages 162–67.

16. Some Properties of the Hydrogen Arc, C. G. Suits. *Journal of Applied Physics* (New York, N. Y.), volume 10, September 1939, pages 648–50.

17. Current Densities, Lumen Efficiency and Brightness in Argon, Nitrogen, Helium and Hydrogen Arc, C. G. Suits. *Journal of Applied Physics* (New York, N. Y.), volume 10, October 1939, pages 730–32.

18. Method of Test for Cloud and Pour Point. *D-97*, American Society for Testing Materials (Philadelphia, Pa.).

19. Dielectric Strength of Insulating Oil of Petroleum Origin. *D877-46T*, American Society for Testing Materials (Philadelphia, Pa.), 1946.

No Discussion

Dynamic Hysteresis Loops of Several Core Materials Employed in Magnetic Amplifiers

HAROLD W. LORD
FELLOW AIEE

ANY good magnetically soft type of core material is suitable for low- and medium-performance magnetic amplifiers. For these types, the permeability and core loss have only minor effects upon the control characteristic of the magnetic amplifier. If the saturation density is reasonably high and the core loss is reasonably low, then the material probably will be suitable for such amplifiers.

For high-performance magnetic amplifiers the situation is very different. When the positive feedback approaches 100 per cent, as it does in the self-saturated types,[1] the core material used for the amplifier plays a major role in determining the control characteristics of the magnetic amplifiers.

An ideal magnetic material for high-performance magnetic amplifiers would have a hysteresis loop such as that shown in Figure 1(A). However, most magnetic amplifier designers would be happy if they could obtain a core material whose hysteresis loop is rectangular and corresponds to that shown in Figure 1(B). This is the actual d-c hysteresis loop of a specimen core composed of 0.005-inch laminations of 65 Permalloy. But, if he chooses this material for use in a 400-cycle magnetic amplifier, he will be disappointed in its performance. The 400-cycle dynamic hysteresis loop of this specimen is shown in Figure 1(C). The 400-cycle dynamic loop departs considerably from rectangularity and is much wider than the d-c loop.

This example might be classed as an extreme case. However, it does emphasize the need for a study of the dynamic hysteresis loops of various magnetic materials in order to determine their relative suitability for magnetic amplifiers. Some of the results of such a study, carried out in the Research Laboratory of the General Electric Company, will be presented in this paper.

The method used for obtaining the dynamic hysteresis loops shown herein has been described in detail in an earlier paper.[2] Figure 2 shows a block diagram of the excitation circuit and the means for displaying a Lissajous figure representing the dynamic hysteresis loop upon the screen of a cathode-ray tube where it can be recorded photographically. Before presenting the dynamic hysteresis loops of several magnetic materials, it is desirable to review briefly the conditions under which they were recorded and to give some justification for imposing these conditions. In the previous paper[2] it was suggested that the flux wave in the core be essentially sinusoidal. All of the major dynamic hysteresis loops shown here were recorded with flux waves such that the voltage induced in the *B* coil had less than 10-per-cent harmonic distortion. This was done for two reasons. First, it complies with the suggested AIEE standards for testing such cores, and

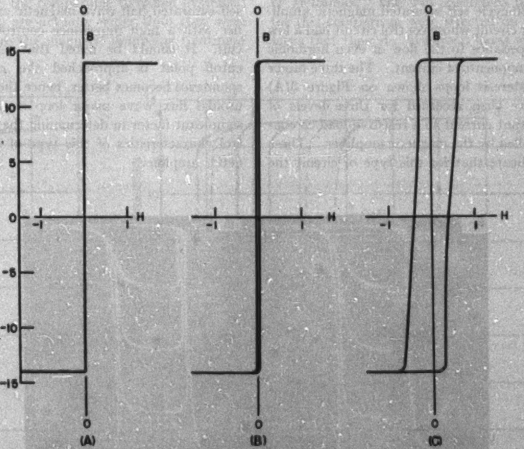


Figure 1(A). Idealized hysteresis loop; (B) d-c hysteresis loop of 0.005-inch tape; (C) dynamic hysteresis loop at 400 cycles of a 0.005-inch tape core of 65 Permalloy

---

Paper 53-117, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted October 6, 1952; made available for printing December 3, 1952.

Harold W. Lord is with the General Electric Company, Schenectady, N. Y.

The author wishes to acknowledge the contributions of D. R. Kearns and D. Lieby, who performed the tests required to obtain the dynamic hysteresis loops included in this paper.

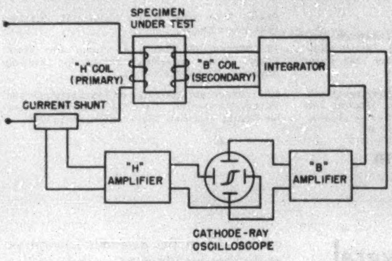
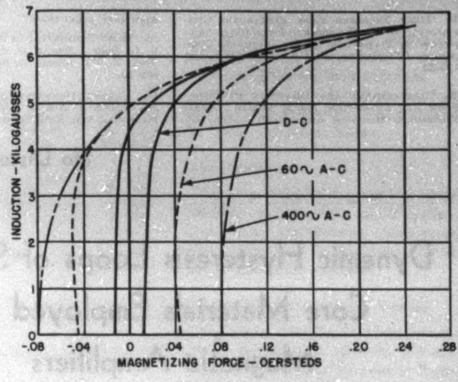


Figure 2 (above). Block diagram of circuit for displaying dynamic hysteresis loops

Figure 4 (right). Hysteresis loops of 0.002-inch Supermalloy



second, our experience indicates that such dynamic hysteresis loops provide a means for evaluating their usefulness in high-performance magnetic amplifiers.

A detailed description of the manner in which the shape and width of the dynamic hysteresis loops influence the magnetic amplifier characteristics is beyond the scope of this paper. However, the composite oscillogram of Figure 3(A) will serve to indicate the value of such data. The 60-cycle major loop shown was recorded with an essentially sinusoidal flux wave. The peak value of magnetomotive force was approximately 20 times the coercive force and hence deflected the beam beyond the limits of the screen. After recording this major loop, the same core specimen was incorporated as one of the inductor elements in a 60-cycle self-saturated magnetic amplifier circuit whose control circuit had a low impedance to the flow of even harmonic components of current. The three minor hysteresis loops shown on Figure 3(A) were then recorded for three levels of output current to a resistive load as controlled by the magnetic amplifier. These indicate that for this type of circuit the minor dynamic hysteresis loops traced over in actual operation coincide with the sinusoidal flux wave major loop in portions of the second and third quadrants. Since these are the portions traced over during the period in which the level of control flux is being established or reset for the next cycle, it is reasonable to conclude that the major dynamic hysteresis loop will play a principal role in the control characteristic and performance of this type of magnetic amplifier.

The correspondence between the minor operating loops and the major dynamic loop in other types of magnetic amplifiers is not this good over the whole control range. Figure 3(B) shows a similar composite oscillogram for a 60-cycle self-saturated half-wave magnetic amplifier with a high impedance control circuit. It should be noted that as the cutoff point is approached the correspondence becomes better, hence the sinusoidal flux wave major loop is still a significant factor in determining the control characteristics of this type of magnetic amplifier.

Comparisons of several oscillograms with each other and with plotted curves are often difficult. Data taken from two or more oscillograms have therefore been plotted to suitable scales on a single set of co-ordinates to provide a more convenient means for comparing them. Data shown in the following figures were taken on spirally wound tape cores with toroidally wound coils.

Figure 4 shows the hysteresis loops for 0.002-inch Supermalloy. This is a high nickel content, unoriented alloy characterized by high values of maximum permeability, a rounded knee, and relatively low values of saturation flux densities. The loops are extremely narrow. This alloy is particularly suited for use in magnetic amplifiers for amplifying very low levels of signal power. (Note: The portions of the dynamic curves above the knee are probably in error since the d-c loop normally should fall inside or coincide with the a-c loops in this region. Also, observe that both scales of this figure are different from those of the following figures.)

Figure 5 shows hysteresis loops for 0.002-inch highly oriented silicon steel. This material is suited to medium and large power handling magnetic amplifiers. In comparable thicknesses it costs less than any of the other alloys shown herein, is more available, and does not use critical materials. Since it has hysteresis loops which are wider (for comparable thicknesses) than those of the other alloys shown, magnetic amplifiers constructed of this material will have relatively poorer performance.

Figure 6 shows the hysteresis loops for 0.002-inch 65 Permalloy. Note particularly the very narrow d-c hysteresis loop, yet the dynamic loops are com-

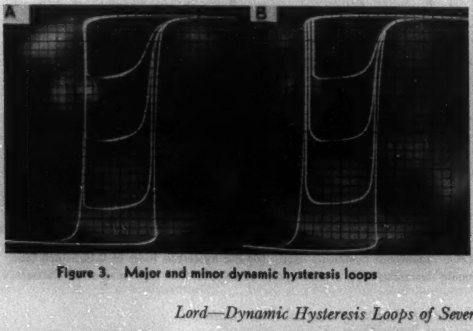


Figure 3. Major and minor dynamic hysteresis loops

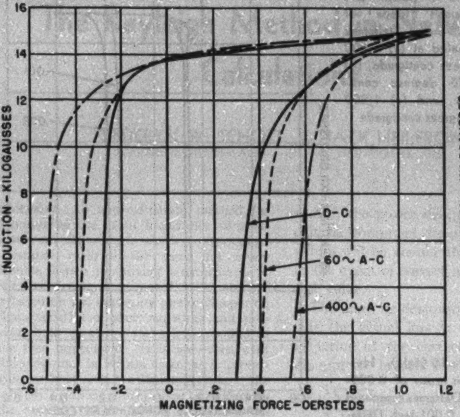


Figure 5. Hysteresis loops of 0.002-inch oriented silicon steel

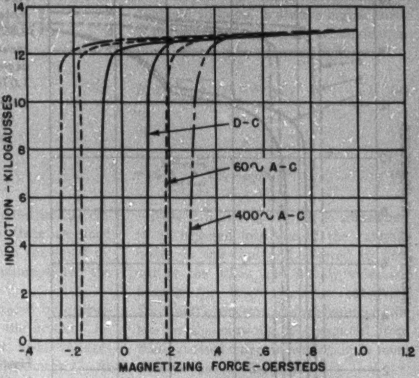


Figure 7. Hysteresis loops of 0.002-inch Deltamax

parable in width to those for Deltamax as shown by Figure 7. The so-called hourglass or wasp-waisted shape shown by the a-c loops of Figure 6 has been observed to occur with highly oriented alloys. This tendency seems to be particularly pronounced when they are domain-oriented alloys. Orientation in 65 Permalloy is obtained by a magnetic anneal so it is a domain-oriented material. At present, 65 Permalloy is not available commercially. In a magnetic amplifier it yields no better performance than that of Deltamax and its saturation density is a little lower than that of Deltamax.

Figure 7 shows the hysteresis loops for a specimen of Deltamax which has been annealed at a temperature slightly above that which produces optimum rectangularity of the d-c loop. However, the dynamic loops shown are more rectangular, so this specimen represents about the best over-all characteristic for this material for use in high-performance magnetic amplifiers.

Figure 8 shows the results obtained when the original specimen used for the Figure 7 data was given a magnetic anneal. This treatment provided a material whose d-c hysteresis loop is not only narrower but is more rectangular. However, the comparable dynamic hysteresis loops are actually wider than they were before the magnetic annealing treatment and tests in a magnetic amplifier showed a corresponding deterioration of performance.

The effect of annealing temperature upon the shape and width of the d-c hysteresis loops is shown by Figure 9. These three specimens were all from the same laboratory produced ingot. Dynamic hysteresis loops have been omitted in this comparison. The amount of widening was roughly the same for the two specimens annealed at the lower temperatures and about 40 per cent less for the one annealed at 1,200 degrees centigrade.

Figures 10 and 11 show the effect of gauge upon the dynamic hysteresis loops of Deltamax. The two specimens used for this comparison were selected on the basis that their d-c hysteresis loops were

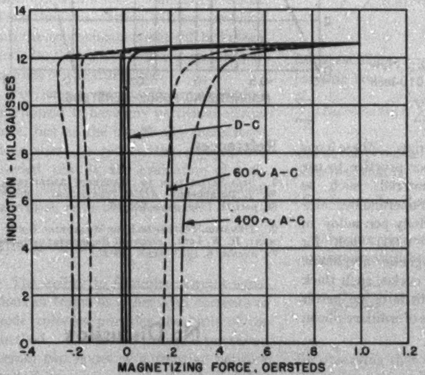


Figure 6. Hysteresis loops of 0.002-inch 65 Permalloy

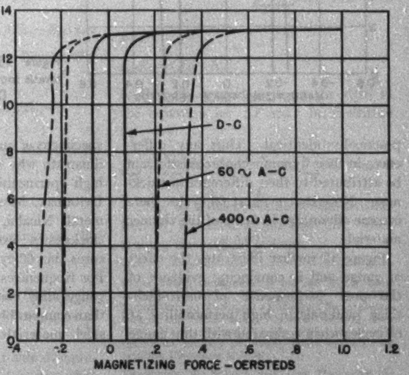


Figure 8. Hysteresis loops of 0.002-inch Deltamax with magnetic anneal

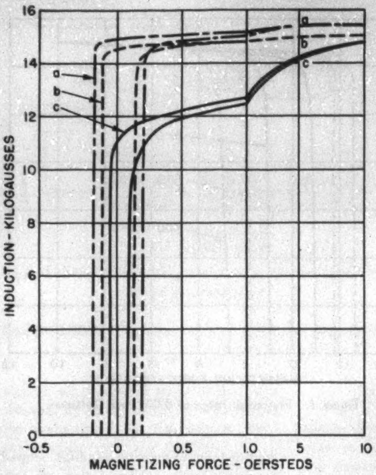
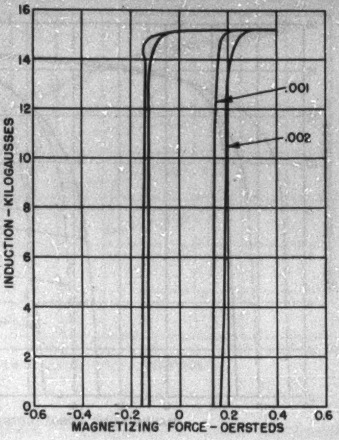


Figure 9 (left). D-c hysteresis loops of 0.002-inch Deltamax annealed at (a) 1,000 degrees centigrade, (b) 1,075 degrees centigrade, and (c) 1,200 degrees centigrade

Figure 10 (right). Hysteresis loops of 0.002-inch Deltamax compared with 0.001-inch Deltamax at 60 cycles



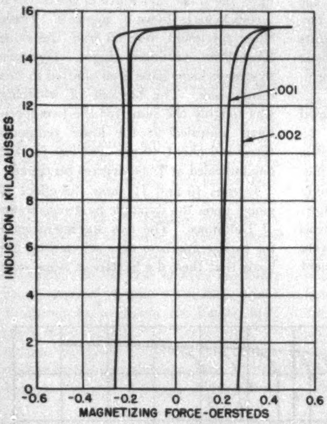
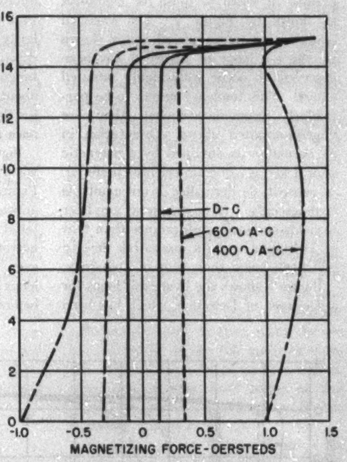


Figure 11 (left). Hysteresis loops of 0.002 inch Deltamax compared with 0.001-inch Deltamax at 400 cycles

Figure 12 (right). Hysteresis loops of 0.010-inch Deltamax



practically identical. Thus any differences in their dynamic characteristics can be attributed to their difference in thickness. Note that even at 60 cycles there is some advantage in favor of the thinner material.

Figure 12 further illustrates the effect of gauge and is convincing evidence of the poor performance of 0.010-inch-thick materials of high permeability at 60 cycles when compared with thin gauge specimens of such materials. These loops illustrate why it is poor practice to use high permeability materials such as Deltamax, Monimax, Supermalloy, Mumetal, Nicaloi, and Molly-permalloy in thicknesses exceeding 0.005 inch for cores in 60-cycle magnetic amplifiers. For frequencies of 400 cycles, such thick gauge alloy cores are little or no better than comparable gauges of ordinary silicon steel materials.

References

1. Self-Saturation in Magnetic Amplifiers, W. J. Dornhoefer. *AIEE Transactions*, volume 68, part II, 1949, pages 835–50.

2. Dynamic Hysteresis Loop Measuring Equipment, H. W. Lord. *Electrical Engineering*, volume 71, number 6, 1952, pages 518–21.

No Discussion

The Rayleigh Method in Network Calculations

FREDERICK W. SCHOTT
ASSOCIATE MEMBER AIEE

JACK HEILFRON
NONMEMBER AIEE

Synopsis: An approximation method for determining the input impedance of a resistive circuit, or one in which the reactance-resistance ratio is the same for every branch, hence permitting treatment as a resistive circuit, is presented. The method is based on the minimum power-dissipation characteristics of such networks and provides both an upper and a lower bound to the true impedance. It is presented with the view that in certain cases, as in power-system short-circuit calculations where the resistive component may often be neglected, the necessity of evaluating high-order determinants or manipulating star-mesh transformations can be eliminated.

LORD Rayleigh seems to have been the first to use an energy method in solving for a characteristic system parameter and his method[1] for approximating the lowest resonant frequency of a mechanical system is well known. This paper presents a method for determining the input impedance of a resistive circuit, or one in which the reactance-resistance ratio is the same for every branch and hence may be treated as a resistive circuit. It is also based on the energy (or power) relationships and provides an approximate solution; hence Rayleigh's name has been applied to the procedure.

The primary contribution of this development is the use of the dual formulation of the circuit equations to provide both an upper and a lower limit to the result. Thus the true solution will not be found but the approximate result will be known, and also that it lies within a certain percentage of the true value.

For simple circuits the method does not appear to have any particular advantage, but as the complexity increases it does seem to reduce the computation needed, that is, the evaluation of high-order determinants or the equivalent star-mesh reduction procedure.

The Method

The ability to formulate circuit equations in terms of either loop currents or node voltages provides the basis of the method. If a known voltage is impressed across two nodes of a resistive network and if the voltages at all other nodes, hereafter called internal nodes, are assumed, the power dissipated in the circuit as computed from these assumed values will be greater than (or equal to, in the event of correct assumptions) the true value.

This may be demonstrated by expressing the actual loss in the system, P_o, in terms of the correct node voltages E_{io} and the branch conductances G_{ij} as

$$P_o=\frac{1}{2}\sum_i\sum_j(E_{io}-E_{jo})^2G_{ij} \tag{1}$$

and comparing this with the dissipation based on the estimated voltages. The power loss calculated from these assumed node voltages E_i may be written as

$$P=\frac{1}{2}\sum_i\sum_j(E_i-E_j)^2G_{ij} \tag{2}$$

The difference ΔP between P and P_o is given by

$$\Delta P=P-P_o=\frac{1}{2}\sum_i\sum_j[(E_i-E_j)^2-(E_{io}-E_{jo})^2]G_{ij}$$

By expanding the quantity in brackets and adding and subtracting appropriate terms, the expression may be rearranged to read

$$\Delta P=\frac{1}{2}\sum_i\sum_j[(E_i-E_{io})-(E_j-E_{jo})]^2G_{ij}+\sum_i\sum_j[(E_i-E_{io})(E_{io}-E_{jo})-(E_j-E_{jo})(E_{io}-E_{jo})]G_{ij}$$

But at each internal node, from Kirchhoff's junction law

$$\sum_j(E_{io}-E_{jo})G_{ij}=\sum_i(E_{io}-E_{jo})G_{ij}=0$$

and at the two external nodes, the voltage is known, that is

$$(E_i-E_{io})=(E_j-E_{jo})=0$$

Therefore the second portion of the summation vanishes and

$$\Delta P=\frac{1}{2}\sum_i\sum_j[(E_i-E_{io})-(E_j-E_{jo})]^2G_{ij} \tag{3}$$

By observing that the right side of equation 3 is positive-definite, it is seen that the power loss, when calculated from the assumed voltages, is always higher than, or in the limit, equal to, the actual dissipation. From this, one may write

$$P_o\leqq\frac{1}{2}\sum_i\sum_j(E_i-E_j)^2G_{ij} \tag{4}$$

or

$$P_o\leqq\sum_{i=1}^{n}\sum_{j=1}^{i}(E_i-E_j)^2G_{ij} \tag{5}$$

Now if a known current is considered to be injected into one of the external nodes in question and allowed to flow out at the other, and if only values of branch (or loop) currents satisfying Kirchhoff's junction law are assumed, the loss in the network as computed from these guessed currents, will again be more than the actual dissipation.

This fact has long been known[2] and may be shown by expressing P_o in terms of the correct branch currents I_{ijo} and the branch resistances R_{ij} as

$$P_o=\frac{1}{2}\sum_i\sum_j I_{ijo}^2R_{ij} \tag{6}$$

The power calculated from the assumed branch currents I_{ij} is, similarly

$$P=\frac{1}{2}\sum_i\sum_j I_{ij}^2R_{ij} \tag{7}$$

and the difference ΔP between P and P_o is

$$\Delta P=P-P_o=\frac{1}{2}\sum_i\sum_j(I_{ij}^2-I_{ijo}^2)R_{ij}$$

By adding and subtracting appropriate terms

$$\Delta P=\frac{1}{3}\sum_i\sum_j(I_{ij}-I_{ijo})^2R_{ij}+\sum_i\sum_j(I_{ij}-I_{ijo})I_{ijo}R_{ij}$$

But

$$I_{ijo}R_{ij}=(E_{io}-E_{jo})$$

so that the second term on the right in the equation for ΔP may be rewritten as

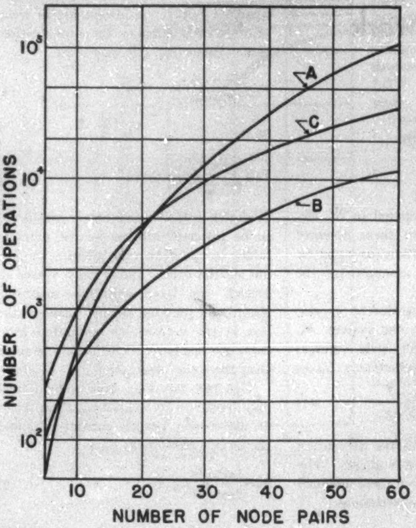
$$\sum_i E_{io}\sum_j(I_{ij}-I_{ijo})-\sum_j E_{jo}\sum_i(I_{ij}-I_{ijo})$$

Also, as a direct result of the satisfaction of Kirchhoff's junction law, when making the original guesses

---

Paper 53-62, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted January 21, 1952; made available for printing December 4, 1952.

FREDERICK W. SCHOTT and JACK HEILFRON are with the University of California, Los Angeles, Calif.



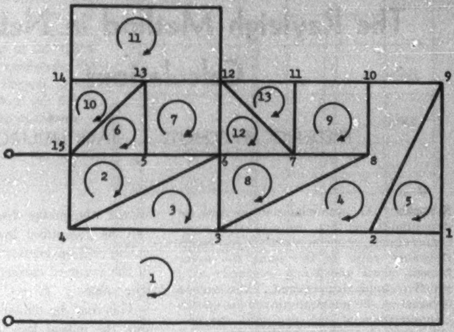


Figure 1 (left). Number of operations required to determine input resistance in the most complex circuit versus number of node pairs using

(A) Network reduction
(B) Energy method without relaxation
(C) Energy method with one relaxation step on the node voltages and a re-estimate of branch currents

Figure 2 (above). Circuit used in the example of Rayleigh's method

$$\sum_{i \text{ or } j} I_{ij} = \sum_{i \text{ or } j} I_{ijo} = 0$$

at all internal nodes and equals the injected current at both external nodes. Therefore

$$\Delta P = \frac{1}{2}\sum_i \sum_j (I_{ij} - I_{ijo})^2 R_{ij} \tag{8}$$

or

$$\Delta P = \sum_{i=1}^{n}\sum_{j=1}^{i} (I_{ij} - I_{ijo})^2 R_{ij} \tag{9}$$

Again, observing the positive-definite nature of the right hand term in equation 9, one may write

$$P_o \leq \sum_{i=1}^{n}\sum_{j=1}^{i} (I_{ij})^2 R_{ij} \tag{10}$$

The actual power dissipation may also be written in terms of the impressed voltage E_I or the injected current I_I and the input conductance G_I or input resistance R_I as

$$P_o = E_I^2 G_I \qquad \text{or} \qquad P_o = I_I^2 R_I$$

These reduce, in the case of unit impressed voltage and unit injected current, to

$$P_o = G_I \qquad \text{and} \qquad P_o = R_I$$

but

$$R_I = \frac{1}{G_I}$$

so, from equations 5 and 10, it follows that

$$\frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{i}(E_i - E_j)^2 G_{ij}} \leq R_I \leq \sum_{i=1}^{n}\sum_{j=1}^{i} I_{ij}^2 R_{ij} \tag{11}$$

Therefore, by computing power loss from the assumed voltages and currents one may obtain an upper and lower bound for the desired input resistance. If the percentage difference between these two limits is less than twice the required accuracy, their average will represent an acceptable result.

On the other hand, if they differ too greatly, a form of the relaxation method developed by Southwell[3,5] may be used to obtain new voltage and current values which will confine the range of equation 11 to within acceptable limits. A modification of the general relaxation procedure by which this reiteration process may be carried out is given in the next section.

It should be mentioned here that in the case of planar networks, instead of using branch currents to obtain an upper limit, the inverse of the circuit may be found and the node voltage analysis used again.

A Relaxation Technique for Improving Results

To reduce the spread of the limits of the computed input resistance to twice the required accuracy, the original guesses may be modified so as to converge upon the correct solution. This is accomplished by the use of relaxation methods.

Relaxation methods[3] begin with an assumed solution, or set of solutions, to a problem and progress by successively correcting the assumptions. Each correction is made so as to eliminate at a point a quantity called the residual, which is determined by the values in the region of that point. Each relaxation step brings the values closer to the overall solution; thus the number of steps

Table I. Resistance and Conductance Chart

Branch Number	Resistance, Ohms	Conductance, Mhos
15-14	1.00	1.00
15-13	0.50	2.00
15- 5	10.00	0.10
15- 4	5.00	0.20
14-13	2.00	0.50
14-12	0.40	2.50
13-12	0.10	10.00
13- 5	4.00	0.25
12-11	0.25	4.00
12- 7	1.00	1.00
12- 6	5.00	0.20
11-10	2.00	0.50
11- 7	0.40	2.50
10- 9	0.25	4.00
10- 8	0.10	10.00
9- 2	10.00	0.10
9- 1	2.00	0.50
8- 7	2.50	0.40
8- 3	4.00	0.25
7- 6	1.00	1.00
6- 5	0.40	2.50
6- 4	0.20	5.00
6- 3	5.00	0.20
4- 3	0.10	10.00
3- 2	2.00	0.50
2- 1	1.00	1.00

required depends upon the correctness of the initial assumptions as well as upon the accuracy desired.

Because the complete network solution is not required, a slight modification of the general relaxation method may be used. In Southwell's procedure, applied to the computation based on the node voltages, the largest residual current at a node in the circuit is eliminated, while in the process suggested herein the residual currents at the internal nodes are eliminated in order, depending on the numbering of the nodes. This makes the determination of the greatest residual current unnecessary, a step which would require additional computation because of the special nature of the problem to be solved. However, the suggested procedure will not converge as rapidly.

The power computed from the assumed voltages may be written as

$$P=\sum_i E_i^2\sum_j G_{ij}-\sum_i E_i\sum_j E_jG_{ij} \qquad (12)$$

and the residual current I_R at node k, which is a measure of the incorrectness of the assumed voltages, may be expressed as

$$I_{Rk}=E_k\sum_j G_{kj}-\sum_j E_jG_{kj}$$

where $j\neq k$, whence

$$\frac{I_{Rk}^2}{\sum_j G_{kj}}=E_k^2\sum_j G_{kj}-2E_k\sum_j E_jG_{kj}+\frac{\left(\sum_j E_jG_{kj}\right)^2}{\sum_j G_{kj}} \qquad (13)$$

Expanding equation 12 by separating that portion of the power due to the voltage at node k results in

$$P=\left(E_k^2\sum_j G_{kj}-2E_k\sum_j E_jG_{kj}\right)+\sum_i E_i^2G_{ik}+\sum_i E_i^2\sum_j G_{ij}-\sum_i E_i\sum_j E_jG_{ij} \qquad (14)$$

where $i\neq j$, $i\neq k$, and $j\neq k$.

By substituting equation 13 into 14, P is given by

$$P=\frac{I_{Rk}^2}{\sum_j G_{kj}}-\frac{\left(\sum_j E_jG_{kj}\right)^2}{\sum_j G_{kj}}+\sum_i E_i^2\sum_j G_{ij}-\sum_i E_i\sum_j E_jG_{ij}+\sum_i E_i^2G_{ik} \qquad (15)$$

If E_k is modified in such a manner that

$$E_k'=\frac{\sum_j E_jG_{kj}}{\sum_j G_{kj}} \qquad (16)$$

then the residual current at node k will be eliminated or

$$I_{Rk}=0$$

So if the new power P' is calculated using E_k', equation 15 will be reduced by the term involving I_{Rk}. Therefore

$$P-P'=\frac{I_{Rk}^2}{\sum_j G_{kj}} \qquad (17)$$

meaning that the dissipation has been reduced by this amount.

If $P-P'=0$, then $I_{Rk}=0$, since all the branch conductances are positive. If the power does not decrease at all, for a systematic relaxation of every internal node voltage, then all of the residual currents in the circuit must have been zero. This is precisely the condition of equilibrium that must be satisfied for the power to be a true minimum, since for this case both of Kirchhoff's laws are satisfied.

A similar process may be carried out with respect to the loop currents. The power loss, as calculated on the current basis, converges to the actual loss in the network if new values of loop cur-

Table II. Computation Chart—Voltage Basis

Node Number	Branch Number	Branch G	Assumed Node E	Branch E	Branch P	E_jG_{ij}	Relaxed Node E	Branch E	Branch P
15	15–14	1.00	1.00	0.10	0.010		1.00	0.15	0.023
	15–13	2.00		0.20	0.080			0.17	0.060
	15– 5	0.10		0.50	0.025			0.43	0.019
	15– 4	0.20		0.50	0.50			0.49	0.048
14	14–15	1.00	0.90			1.00	0.85		
	14–13	0.50		0.10	0.005	0.40		0.023	0.003
	14–12	2.50		0.10	0.025	2.00		0.060	0.010
		4.00				3.40			
13	13–15	2.00	0.80			2.00	0.83		
	13–14	0.50				0.43			
	13–12	10.00		0.00	0.00	8.00		0.037	0.014
	13– 5	0.25		0.30	0.023	0.12		0.26	0.017
		12.75				10.55			
12	12–14	2.50	0.80			2.18	0.79		
	12–13	10.00				8.27			
	12–11	4.00		0.10	0.04	2.80		0.085	0.029
	12– 7	1.00		0.20	0.04	0.60		0.122	0.015
	12– 6	0.20		0.30	0.018	0.10		0.265	0.014
		17.70				13.95			
11	11–12	4.00	0.70			3.16	0.70		
	11–10	0.50		0.15	0.011	0.28		0.165	0.014
	11– 7	2.50		0.10	0.025	1.50		0.037	0.003
		7.00				4.94			
10	10–11	0.50	0.55			0.35	0.54		
	10– 9	4.00		0.05	0.01	2.00		0.065	0.017
	10– 8	10.00		0.00	0.00	5.50		0.001	0.000
		14.50				7.85			
9	9–10	4.00	0.50			2.16	0.47		
	9– 2	0.10		0.30	0.009	0.02		0.292	0.008
	9– 1	0.50		0.50	0.125	0.00		0.475	0.113
		4.60				2.18			
8	8–10	10.00	0.55			5.40	0.54		
	8– 7	0.40		0.05	0.001	0.24		0.127	0.006
	8– 3	0.25		0.05	0.000	0.12		0.051	0.000
		10.65				5.76			
7	7–12	1.00	0.60			0.79	0.67		
	7–11	2.50				1.70			
	7– 8	0.40				0.21			
	7– 6	1.00		0.10	0.001	0.50		0.143	0.020
		4.90				3.26			
6	6–12	0.20	0.50			0.16	0.53		
	6– 7	1.00				0.67			
	6– 5	2.50		0.00	0.00	1.25		0.042	0.004
	6– 4	5.00		0.00	0.00	2.50		0.015	0.000
	6– 3	0.20		0.00	0.00	0.10		0.035	0.000
		8.90				4.68			
5	5–15	0.10	0.50			0.10	0.57		
	5–13	0.25				0.20			
	5– 6	2.50				1.31			
		2.85				1.61			
4	4–15	0.20	0.50			0.20	0.51		
	4– 6	0.50				0.26			
	4– 3	10.00		0.00	0.00	5.00		0.02	0.004
		10.70				5.46			
3	3– 8	0.25	0.50			0.13	0.49		
	3– 6	0.20				0.11			
	3– 4	10.00				5.00			
	3– 2	0.50		0.30	0.045	0.10		0.307	0.047
		10.95				5.34			
2	2– 9	0.10	0.20			0.05	0.18		
	2– 3	0.50				0.24			
	2– 1	1.00		0.20	0.04	0.00		0.183	0.030
		1.60				0.29			
					0.582				0.518

Table III. Computation Chart (Current Basis)

Loop	Branch Number	Branch R	Assumed Branch I	Branch P	I_j's*	Assumed† Loop I	I_jR_{ij}	Relaxed Loop I	Branch I	Branch P
1	15–4	5.0	0.10	0.05		1.0			0.136	0.093
	4–3	0.1	0.20	0.004					0.195	0.004
	3–2	2.0	0.30	0.18					0.310	0.194
	2–1	1.0	0.40	0.16					0.393	0.155
2	15–5	10.0	0.10	0.10	2–6	0.9	8.00	0.86	0.038	0.015
	5–6	0.4	0.10	0.004	2–7		0.32		0.054	0.000
	6–4	0.2	0.10	0.002	2–3		0.16		0.059	0.000
	4–15	5.0			2–1		5.00			
		15.6					13.48			
3	6–3	5.0	0.00	0.00	3–8	0.8	4.00	0.80	0.098	0.048
	3–4	0.1			3–1		0.10			
	4–6	0.2			3–2		0.17			
		5.3					4.27			
4	10–9	0.25	0.70	0.123	4	0.7	0.00	0.69	0.69	0.120
	9–2	10.00	0.10	0.10	4–5		6.00		0.083	0.069
	2–3	2.0			4–1		2.00			
	3–8	4.0	0.20	0.004	4–8		3.20		0.017	0.001
	8–10	0.10	0.20	0.004	4–9		0.05		0.268	0.007
		16.35					11.25			
5	9–1	2.0	0.60	0.72	5	0.6	0.00	0.60	0.607	0.740
	1–2	1.0			5–1		1.00			
	2–9	10.0			5–4		6.90			
		13.0					7.90			
6	15–13	0.5	0.50	0.125	6–10	0.8	0.15	0.83	0.536	0.144
	13–5	4.0	0.00	0.00	6–7		3.20		0.016	0.001
	5–15	10.0			6–2		8.64			
		14.5					11.99			
7	13–12	0.1	0.50	0.025	7–11	0.8	0.03	0.81	0.528	0.028
	12–6	5.0	0.00	0.00	7–12		4.00		0.030	0.004
	6–5	0.4			7–2		0.35			
	5–13	4.0			7–6		3.30			
		9.5					7.68			
8	8–3	4.0			8–4	0.8	2.76	0.70		
	3–6	5.0			8–3		4.03			
	6–7	1.0	0.00	0.00	8–12		0.80		0.073	0.009
	7–8	2.5	0.30	0.225	8–9		1.25		0.285	0.204
		12.5					8.84			
9	11–10	2.0	0.50	0.50	9	0.5		0.42	0.422	0.360
	10–8	0.1			9–4		0.07			
	8–7	2.5			9–8		1.76			
	7–11	0.4	0.20	0.02	9–13		0.28		0.093	0.000
		5.0					2.11			
10	15–14	1.0	0.30	0.09	10	0.3	0.00	0.29	0.290	0.085
	14–13	2.0	0.00	0.00	10–11		0.60		0.008	0.000
	13–15	0.5			10–6		0.41			
		3.5					1.01			
11	13–14	2.0			11–10	0.3	0.58	0.28		
	14–12	0.4	0.30	0.04	11		0.00		0.282	0.032
	12–13	0.1			11–7		0.08			
		2.5					0.64			
12	12–7	1.0	0.10	0.01	12–13	0.8	0.70	0.78	0.275	0.076
	7–6	1.0			12–8		0.70			
	6–12	5.0			12–7		4.05			
		7.0					5.45			
13	12–11	0.25	0.70	0.12	13	0.7	0.00	0.52	0.52	0.066
	11–7	0.40			13–9		0.17			
	7–12	1.0			13–12		0.78			
		1.65					0.95			
				2.64						2.45

*2–6 indicates loop current I_2 − (loop current)I_6 = branch current I_{15-5}.

†This column arrangement is used since, without relaxation, calculations end with column 5.

rents are given by

$$I_k' = \frac{\sum_{j} I_j R_{kj}}{R_k}$$

where

R_k = total resistance of loop k

R_{kj} = resistance common to loop k and loop j

Some difficulty arises due to the need for branch currents to compute power and loop currents to use the iteration method presented in the preceding. Since there is no need for determining loop currents unless relaxation is found to be necessary, it is recommended that branch currents be assumed initially. Then if reiteration is needed, loop currents can be computed readily.

Comparison of the Rayleigh Method with Other Methods

The amount of computation involved will depend, in general, on the type and complexity of the network under consideration. For the most complex case, wherein every node is directly connected to every other node by one element (a circuit rarely encountered in practice), the exact method most applicable for solution of the problem is that of network reduction. This, in effect, reduces the labor of expanding high-order determinants (if the number of nodes and branches is large), but still requires considerable effort. Richardson[4] has given systematic procedures for carrying out this reduction.

If a unit operation is defined as the addition, subtraction, multiplication, or division of two numbers, it is possible in certain cases to determine the number of operations required to obtain a value for the input impedance at any point in the circuit.

Figure 1 illustrates the number of operations necessary for solution versus the number of node pairs in the system, for both the network reduction process and the method developed herein. It is seen that, for a large number of nodes, a great saving results from the use of our method providing the original guesses are accurate enough so as to require no relaxation. Figure 1 deals with the most complex case as described here and is based on the following expressions

$$O_r = \frac{7}{2}(N_p^2 + N_p) - (N_p - 1)$$

$$O_n = \frac{N_p}{2}[N_p^2 + N_p - 2]$$

$$O_v = 3N_p^2 - 4N_p + 1$$

where

O_r = operations required to compute input resistance by averaging the limits obtained from assumed voltages and currents (without relaxation)

O_v = operations required to recalculate one entire set of node voltages by relaxation procedures

O_n = operations required to find input resistance using network reduction, for the most complex case

N_p = number of node pairs in circuit

In the event iteration is necessary, for complex cases where the number of node equations is far less than the number of loop equations, it is desirable to use the relaxation process on the node voltages only, and then, on the basis of the modified potentials, re-estimate the branch currents. If this is done, no loop currents need be found at all.

In a less complicated circuit, especially if the number of loops is of the same order of magnitude as the node pairs, convergence upon the correct input resistance can best be carried out by relaxing the residuals in both sets of guessed quantities.

Example

An example is shown in Figure 2. The values of resistance and conductance of the branches have been chosen arbitrarily and are not intended to represent any specific physical problem but rather to illustrate the method; they are indicated in Table I.

The calculations have been carried out in the tabular form; see Tables II and III. For the calculation on the node voltage basis, the use of the first six columns is obvious, the sum of the values in column 6 giving the first estimate of the input conductance. Column 7 and the subtotals shown are needed for the relaxation operation to derive the modified node voltages according to equation 16. Each subgroup of column 7 is computed alternately with the corresponding element of column 8, this process being carried out in order. The new values of column 8 are then used in lieu of those of column 4 in all subsequent calculations in column 7. Columns 9 and 10 are similar to columns 5 and 6. If an additional relaxation step were performed another set of columns similar to columns 7, 8, 9, and 10 would be needed.

In computing the total power dissipated, that dissipated in each branch must be included but once; therefore, it may be noted that numerous blank spaces occur in columns 5, 6, 9, and 10.

The column arrangement used for making calculations on the current basis is similar.

From Table II it can be seen, on the basis of the voltage assumptions, that:

$R_I=1/(0.582)=1.72$ ohms (from original values)

$R_I=1/(0.518)=1.93$ ohms (after one relaxation operation)

From Table III we find that:

R_I on the basis of original guesses equals 2.64 ohms

R_I on the basis of relaxed currents equals 2.45 ohms

Results

The average value of R_I as computed from original estimates is 2.18 ohms. The maximum possible error in this result is 21.2 per cent.

The average value for R_I as computed from relaxed quantities is 2.19 ohms. The maximum possible error in this result is 11.8 per cent. Therefore, with one relaxation step, the range of possible values for input resistance has been halved, although the average, or mid-range, value has not changed appreciably.

The correct value found from a network-analyzer solution of the problem was found to be 2.22 ohms; therefore the actual error after one relaxation operation is

$$\frac{2.22-2.19}{2.22}(100)=1.35 \text{ per cent}$$

It has been observed in applying these methods that there seems to be an equal-error tendency, that is, the upper and lower bounds tend to deviate from the correct value by roughly the same magnitude; hence the average tends to approximate the true result reasonably well.

References

1. Theory of Sound (book), **Lord Rayleigh.** Dover Publications, New York, N. Y., first American edition, volume I, 1945.

2. Elements of the Mathematical Theory of Electricity and Magnetism (book), **Sir J. J. Thomson.** Cambridge University Press, Cambridge, England, 1895.

3. Relaxation Methods in Engineering Science (book), **Sir R. V. Southwell.** Oxford University Press, London, England, 1940.

4. Electrical Network Calculations (book), **D. E. Richardson.** D. Van Nostrand Company, Inc., New York, N. Y., 1946.

5. Network Analysis by the Chain-Relaxation Method, **L. Tasny-Tschiassny.** *Journal*, Institution of Electrical Engineers (London, England), volume 95, part III, number 35, May 1948.

No Discussion

Fluorine-Containing Gaseous Dielectrics

G. CAMILLI
FELLOW AIEE

R. E. PLUMP
NONMEMBER AIEE

Synopsis: The characteristics of gaseous chemical compounds containing fluorine are reviewed. Relative dielectric strengths, toxicities, space charge effects, electrode material, pressure effects, and so forth, are discussed. Several explanations accounting for observed differences in dielectric strength among gases are noted and shown to be inadequate for the complete understanding of all cases. A line of attack for fuller understanding is suggested in the considerations of the molecular energy systems. Some possible potentialities of gaseous fluorine compounds in electric equipment are also reviewed.

GASEOUS fluorocompounds are specifically of interest in high-voltage applications because of their generally inert character and high dielectric strength. Some of these compounds such as the Freons have already found wide use in refrigeration and air conditioning. Hydrogen fluoride is used in the petroleum industry as a catalyst for the production of high-octane gasoline.

Fluorine chemistry[1] has attained a rather unique position due to the fact that the element fluorine is the most active chemical we have, attacking nearly all other substances with the peculiar result that many of the combinations are then among the most inert substances. For example, fluorine attacks sulfur and the product is chemically, thermally, and electrically stable. Technological progress in the production and handling of fluorine and widespread study of fluorine compounds have made a separate field in general chemistry. Before the last war, fluorine was made in only gram quantities, and fluorocarbons were laboratory curiosities. Today both the variety and quantity of available fluorocompounds is large. It is of interest to review some of the gases for electrical use because of possible advantages to be gained.

The fluorogases of most interest are the Freons, sulfur hexafluoride, and the fully fluorinated compounds such as C_3F_8.

Factors which must be considered in the choice of an insulating gas for electrical use are:

1. High dielectric strength.
2. Low temperature of condensation.
3. Chemical inertness, nontoxicity, and nonflammability.
4. Good heat conductivity.
5. Availability and cost.

Other properties such as high density can

Paper 53-126, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted June 13, 1952; made available for printing December 5, 1952.

G. Camilli and R. E. Plump are both with the General Electric Company, Pittsfield, Mass.

The authors wish to express their appreciation to Dr. J. D. Cobine for reading the manuscript and making valuable suggestions.

Table I. Some Properties of Various Fluorocompounds

Formula	Trade or Chemical Name	Boil Point, °C	Relative Dielectric Strength*	Toxicity†	Molecular Weight	Grams per Centimeter at S.T.P.‡	Specific Heat, Calories per Gram at 30°	Crit. Temp., °C	Crit. Pres. Atmos.	Freeze Point, °C
CHF_3	fluoroform Freon-23	−82.5	0.75		70	28.7×10^{-4}	0.15	25	50	−160
$CHClF_2$	Freon-22	−40.8	1.3	5 5A	86.5	35.8×10^{-4}	0.152	96	48.7	−160
CF_4	Freon-14	−128	1.1	prob. 6	88	36.1×10^{-4}	0.132^{-80°	−47.3	36.9	−184
$CClF_3$	Freon-13	−81.5	1.4	prob. 6	104.5	42.9×10^{-4}	0.138^{-30°	28.8	38.1	−181
CCl_2F_2	Freon-12	−29.8	2.4–2.5	prob. 6	121	49.6×10^{-4}	0.147	112	40.1	−158
CCl_3F	Freon-11	23.8	3–4.5	5A	137.5	(61.5×10^{-4})§	0.135	198	43.2	−111
C_2F_6	hexafluoro ethane	−78	1.8	to be tested	138	56.6×10^{-4}	to be measured	24.3	32.6	−101
C_3F_8	octafluoropropane	−37	2.0	to be tested	188	77.1×10^{-4}	to be measured	70.5	26.5	−160
C_4F_{10}	decafluorobutane	−2.5	CA 2.5 extrapd	to be tested	238	97.5×10^{-4}	to be measured	113	20	−80
C_5F_{12}	dodecafluoropentane	29.3	2.8	to be tested	288	(118×10^{-4})§	to be measured			−125.5
SF_6	sulfur hexafluoride	−63	2.0–2.5	physio inert	146	59.9×10^{-4}	0.143	54	36.7	−50.8 melting point
$CClF=CF_2$	chloro-trifluoro ethylene	−28		4 or 5	116.5	52.2×10^{-4}		107		
$CClF_2.CF_3$	Freon-115	−38	2.8	prob. 6	154.5	69.1×10^{-4}	0.164	80	(32)	−106
$CClF_2.CClF_2$	Freon-114	3.6	2.8	prob. 6	171	76.5×10^{-4}	0.160	145.7	32.1	−94
$CCl_2F.CClF_2$	Freon-113	47.6	2.6 0.4 atmosphere	4 or 5	187	(83.5×10^{-4})§	0.161^{60°	214.1	33.7	−35
$CCl_2F.CCl_2F$	Freon-112	93		4 or 5	204	(91.5×10^{-4})§		278	(34)	26

* Relative to N_2 or air, collected from various sources.

† Mostly from reference 43.

‡ Calculated as ideal gases; nearly true values, but mainly for comparison.

§ Parentheses indicate questionable values.

Most of these data have been obtained from commercial sources including Kinetic Chemicals, Inc., Minnesota Mining and Manufacturing Company, and General Chemical Company.

be advantageous, but in general a compromise among various properties will probably be necessary in any particular application.

Table I shows the physical variations to be found among the dielectric gases.

Dielectric Strength

VARIOUS ELECTRODE CONFIGURATIONS AND CONDITIONS OF VOLTAGE APPLICATION

In 1937 Charlton and Cooper[2] obtained five fluorocompounds among other gases and determined their relative strengths to that of nitrogen. The electrodes were curved, 2-inch radius, nickel surfaces separated 1/8-inch or 1/4-inch gap distance and the voltage applied was direct, in most cases with the gas at atmospheric pressure. The relative values when compared with nitrogen under the same conditions were: CH_2F_2, 0.69; CCl_2F_2(Freon-12), 2.4; CCl_3F(Freon-11), 3.0; $C_2Cl_3F_3$, 2.6; and $C_2Cl_2F_4$, 2.8. These authors also seem to have been the first to study known mixtures of Freon with nitrogen or hydrogen obtaining strengths proportional to the compositions of the mixtures.

Between a point and sphere in a 3-millimeter gap with a direct voltage, Pollock and Cooper[3] found with increasing pressure a certain anomalous behavior when the point was positive in Freon-12 and sulfur hexafluoride. This was similar to an effect also observed in air[4] under high pressure, and this will be discussed a little later. Nevertheless, a selection of data for the positive point at 4 atmospheres pressure gives breakdown values of 13 kv for nitrogen, 42 kv for Freon-12, and 48 kv for sulfur hexafluoride.

A comparison[5] of nitrogen and of Freon-12 with transformer oil was made in both a uniform field and between rods, 60-cycle alternating current and impulse, under different pressures.

The impulse and the 60-cycle crest breakdown strength of nitrogen were substantially the same when tested between spheres at less than radius spacing for atmospheric pressure to 200 pounds absolute; the impulse and 60-cycle crest breakdown at CCl_2F_2 were likewise the same but from atmospheric pressure to 80 pounds absolute. In order to obtain the same impulse strength as oil N_2 required a pressure of 175 pounds per square inch absolute and F-12 needed 62 pounds per square inch absolute. To obtain a 60-cycle strength equal to oil, N_2 required a pressure of 80 pounds per square inch absolute and F-12 a pressure of 25 pounds per square inch absolute. These tests were made between 6.25-centimeter spheres spaced 1 centimeter apart.

Between the ends of 1/2-inch square rods spaced 6 centimeters apart, nitrogen at about 63 pounds absolute or Freon-12 at atmospheric pressure were equal to or better than 10-C oil under 60-cycle testing. Under the same electrode arrangement with impulse testing, the strengths of the gases were less than that of oil, with N_2 up to 200 pounds absolute and Freon-12 up to 80 pounds absolute.

In comparison, with d-c uniform fields[6] in a 0.2-inch gap at 125 pounds pressure absolute, the breakdown gradient for Freon-12 was found to be 1,200 kv per inch and the pressure distance with voltage showed a small deviation from Paschen's law.

Later, Skilling and Brenner,[7] using spheres and 60-cycle alternating current found that for small spacings the strength of Freon-12 followed the relation

$$\text{kilovolt breakdown} = 183PS + 4$$

in which P was the actual atmospheres pressure and S was the gap spacing in inches. For example in a gap of 0.1 inch at atmospheric pressure the strength of Freon was about 23 kv while that of air was 10 kv. At 30 pounds gauge (about 3 atmospheres absolute) the strength of Freon was about 60 kv in the same gap.

The impulse strength of Freon-12 and of sulfur hexafluoride in a uniform field between spheres with a positive wave under 45 pounds absolute pressure had been found to be about the same as that of transformer oil[8] and their strengths

Table II. Linked Toroids, 7-Inch Inside Diameter, 2-Inch Effective Gap Comparison Between Air and SF_6, 60-Cycle and Impulse Positive Wave, 20 and 25 Pounds Absolute, 2-Inch Gap

	Kilovolt Spark-over	
	60-Cycle	Positive Impulse
At 20 pounds absolute		
Air	90	150
SF_6	210	360
At 25 pounds absolute		
Air	110	180
SF_6	250	450

Table III. Comparison of Air with Sulfur Hexafluoride, 3-Inch Plane Disk Electrodes and 1/2-Inch Sphere-to-Plane Disk. Gap in Both, 1/2 Inch; 60-Cycle and Positive Wave Impulse at Three Pressures

	Kilovolt Spark-over	
	60-Cycle	Positive Impulse
3-inch plane disks, atmospheric pressure		
Air	23	
SF_6	52	96
1/2-inch sphere-to-plane disks		
Atmospheric pressure		
Air	16	15
SF_6	42	57
30 pounds absolute		
Air	24	30
SF_6	60	96
45 pounds absolute		
Air	45	50
SF_6	90	135

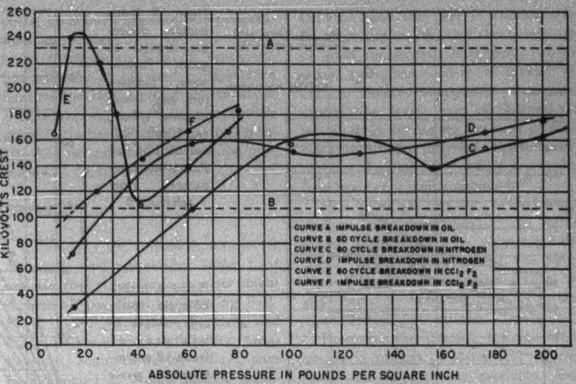


Figure 1. 60-cycle and impulse breakdown between 1/2-inch square rods spaced 6 centimeters in N_2, CCl_2F_2, and 10-C oil. From reference 5

relative to nitrogen were 1.4. The 60-cycle strength of SF_6 was about 75 per cent that of oil, and when the gas pressure was only 30 pounds absolute this factor fell to 63 per cent. Freon-12, however, at 45 pounds absolute had 90 per cent of oil strength and 77 per cent at 30 pounds absolute.

In a nonuniform field, using a pointed tungsten rod and sphere separated 1 inch, the 60-cycle a-c strength of SF_6 was, at 45 pounds absolute, 114 per cent that of oil; but the corresponding value for Freon-12 was only 40 per cent. However, when the Freon-12 was at 30 pounds then the strength was 62 per cent that of oil. This peculiar effect had already been observed by Nonken[5] whose curves showed the beneficial effect of a few pounds pressure above atmospheric but reversal and depreciation at higher pressures in nonuniform fields. In order to produce uniform fields at wider gap spacings, linked toroids were used[8] with an effective gap of 2 inches and with results summarized in Table II. The large difference in dielectric strength between air and sulfur hexafluoride in this case is evident.

Work on SF_6 continued[9] with other gap arrangements, notably with 3-inch plane disks and with a 1/2-inch sphere to 3-inch plane. Typical results as shown in Table III again exemplify the general dielectric advantage of SF_6 over air, in a 1/2-inch gap.

The highest test pressures with uniform field yet reported on SF_6 alone and in mixtures are those of Kusko, who in fact employed such high pressures that the dilution of the SF_6 became necessary in order to obtain breakdowns even with his very high voltage generator. In a 1/2-inch gap, for example, with a uniform d-c field, the following four gases had the following very informative breakdown values in kilovolts: SF_6 800; CO_2, 435; N_2, 350; He, 55, at only 215 pounds per square inch absolute pressure between stainless steel electrodes. Further effects of pressure will be discussed shortly.

Meanwhile, interesting results have been obtained on the series of fluorocarbons $C_nF_{2n} + 2$, from C_1 to C_8[10] and also with perfluoroethyl ether, furane, and bromotrifluoromethane, and others.[11] Some sample results, read from curves, are best summarized in Table IV, all with 60-cycle alternating voltage atmospheric pressure and within the gap spacings stated. The electrode configurations followed those of Camilli and Chapman.[8]

Effect of Pressure

The dielectric strength of a gas is generally raised by increasing pressure. This is most true in the uniform field, as shown by Nonken[5] with spheres, at pressure up to 200 pounds per square inch absolute and there appears to be no decrease in strength at any pressure within practical considerations. In the nonuniform field of a 6-centimeter rod gap, see Figure 1, at 60 cycles, a decrease in spark-over strength of CCl_2F_2 sets in at a pressure of little more than atmospheric. Impulse voltages generally show a rise with pressure although there are exceptions as shown in Figure 2. With direct current there appears to be no critical pressure among the gases reported in a uniform field. Also with direct current with needle-to-plane configuration, if the point is negative, there appears to be no critical pressure. However, when the point is positive, a critical pressure always seems to exist for any gas forming negative ions as similarly shown by Pollock and Cooper[2] and others. As expected, the critical pressure is different for different gases, and the simplest gases, such as oxygen, have higher pressure stability. While this phenomenon of spark-over has been observed repeatedly, the fact that prebreakdown corona voltages rise steadily with pressure under all conditions is noteworthy. The effect of

Table IV. Sample Results on Fluorocompounds, 60-Cycle, Atmospheric Pressure

Electrodes and Spacing	Kilovolt Breakdown, Crest				
From reference 10:	C_4F_{10}	C_3F_8	SF_6	C_2F_6	N_2
1/2-inch radius hemispheres, 0.1-inch gap	28	20	22	15	10
1/2-inch radius hemispheres, 0.25-inch gap		50	52	38	
Point to hemisphere, 0.1-inch gap	20	15	14	13	7
Point to hemisphere, 0.25-inch gap	45 (extrapolated)	37	33	27	10

	Kilovolt Breakdown, Rms							
From reference 11:	$(C_2F_5)_2O$	c-C_4F_8O	C_4F_{10}	C_3F_8	CF_4	CF_3Br	N_2	SF_6
0.2-inch gap, 1-inch diameter spheres	46	42	37	32	15	28		36
0.3-inch gap, 0.1-inch rounded rod to 1-inch sphere	42	38	38	30	12	19	8	24

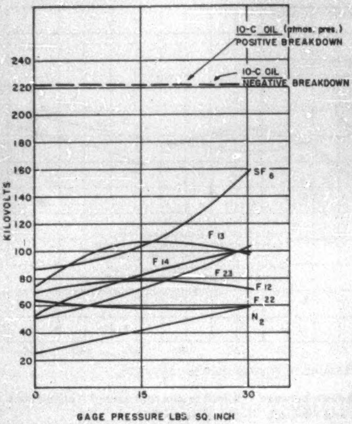


Figure 2 (left). Impulse dielectric strength of gases. Positive wave, tungsten rod, and 1-inch diameter sphere, gap 1 inch. Values of 10-C oil shown for comparison. From reference 8

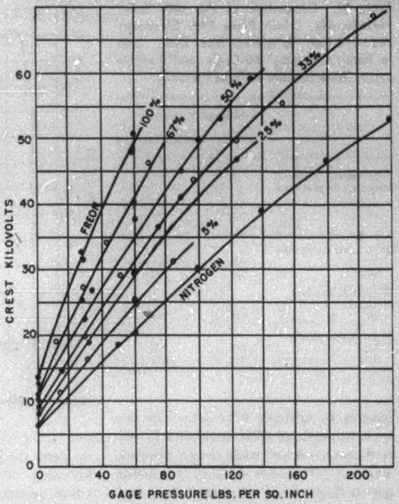


Figure 4 (right). 60-cycle sparking voltages in mixtures of Freon-12 and nitrogen versus pressure. Gap 0.05 inch, spherical electrodes per cent Freon by volume as noted on the curves. From reference 7

pressure on spark-over from a point electrode, by reducing the diffusion of positive space charges left by electron avalanches into the point, is a distortion and concentration of the applied field in the gap, and the gap length is thus effectively shortened by the compressed positive space charge. In other words, when a point is positive, electron avalanches into the point leave behind slow-moving heavy positive ions. When the pressure is low, these diffuse outwardly rather quickly; but when the pressure is increased, the positive ions are compressed and diffusion is reduced, and this is equivalent to a shortened gap.

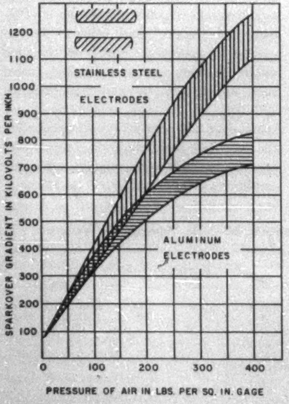


Figure 3. D-c dielectric strength of air between stainless steel or aluminum electrodes. Gap spacings of 1/4, 1/2, and 3/4 inches. From reference 6

When the point is negative, the electron avalanches travel away from the point and the remaining positive ions move slowly toward the point, in effect creating a more uniform field and one that is partially neutralized around the negative point by the positive ions, and this raises the required voltage for breakdown.

Effect of Temperature

In general there is no effect of temperature on the dielectric strength providing that the density of the gas is held constant by being in a firmly closed system. That temperature should enter into certain phenomena such as the critical pressure on positive points, just mentioned, would seem likely since the mobility of the molecules and ions always increases with temperature. This should raise the critical pressure because the rate of diffusion of space charge also increases with temperature.

Effect of Electrode Metals

The literature contains little or nothing on the effect of metals on the dielectric strength of fluorocompounds. However, the corrosion after arcing naturally is different on different metals. For example, aluminum was found more resistant than brass, and steel was even less resistant than either.[7]

While not concerned with fluorogases, it is interesting to note that other authors have found an effect of the electrode metal at subatmospheric pressures, for example, Ehrenkranz[12] and Jacobs and LaRocque.[13] Strangely, at pressures other than in the neighborhood of atmospheric, the effects of particular metals as electrodes appear more pronounced. This is still further illustrated[14] by differences between aluminum and stainless steel on carbon dioxide, nitrogen, and air, the last shown in Figure 3 at very high pressures.

Effect of Foreign Solids and Irregularities on the Electrodes

Cotton lint and dust have been shown to lower the breakdown strength.[15] The effect was most pronounced when the gas contained moisture because the dust absorbed the moisture with a probable increase in dielectric constant and conductivity. Likewise, it has been shown that surface irregularities cause a low breakdown value which rises on continued sparking due to removal of the irregularities. Howell[4] observed this effect with air.

Effect of Mixtures

Charlton and Cooper[2] in 1937 showed that mixtures of Freon-12 with N_2 and with H_2 had approximately a proportionally higher strength with increasing Freon concentration. This result was confirmed and furthermore was shown at small concentrations of F-12 to be disproportional by Skilling and Brenner[7] as shown in Figures 4 and 5. For mixtures greater than 5 per cent in uniform fields, these authors found the linear relationship $V = (88PS + 1.9)(1 + 0.8F)$ in which F is the Freon fraction by volume.

Similar results were obtained by Gossens[15] with mixtures of Freon-12 and air. For instance, 10 per cent of Freon in air at about 8 atmospheres pressure increased the air strength from 110 kv to 150 kv, or about 30 per cent in a uniform 60-cycle field. In a nonuniform field, a maximum is shown in Figure 6 in accordance with the previously discussed space charge effects.

At 27 atmospheres in a uniform d-c field, 8 per cent of SF_6 raised the strength of nitrogen from 430 kv to 680 kv, or by 58 per cent in a 1/2-inch gap. At 13.6 atmospheres the increase in strength amounted to 40 per cent.

One effect of mixtures is to change the temperature of condensation. This follows from Dalton's law of partial pressures which states that[16] in a mixture of gases the total pressure is the sum of the pressures which the separate gases would exert if each occupied the total volume of the mixture at the same temperature. For example, at atmospheric pressure Freon-12 condenses at about −29 degrees centigrade, but if a mixture by volume of 63 per cent Freon-12 and 37 per cent nitrogen at atmospheric pressure is cooled, a temperature of about −40 degrees centigrade should be attainable before condensation of the Freon-12 occurs. The presence of the nitrogen, by diluting the Freon-12, lowers the temperature for the occurrence of a "dew point." At another pressure, for example at 45 pounds absolute, the temperature of condensation of Freon-12 is found to be about 1 degree centigrade but with the 37 per cent nitrogen in a mixture the dew point drops to about −12 degrees centigrade, the temperature at which the partial pressure of the Freon-12 is the same as it would be in the pure state at 1 degree centigrade.

Summary

Considering the various factors so far discussed, and inspecting Table I, it appears that gases for indoor use might include all the Freons (excepting F-11, F-112, and F-113), the C_2F_6, C_3F_8, and C_4F_{10} fluorocarbons, and sulfur hexafluoride. Some, however, will at once be eliminated by having inferior dielectric strength. For outdoor use (−40 degrees low temperature limit) the selection of undiluted gases is reduced to SF_6, C_2F_6, and Freons 23, 22, and 13. Of these, SF_6 has the highest dielectric strength. Nevertheless, in view of the wide possibilities in fluorine chemistry, other compounds in the near future may be produced surpassing SF_6 in over-all qualities, and mixtures with lower temperatures of condensations may be useful.

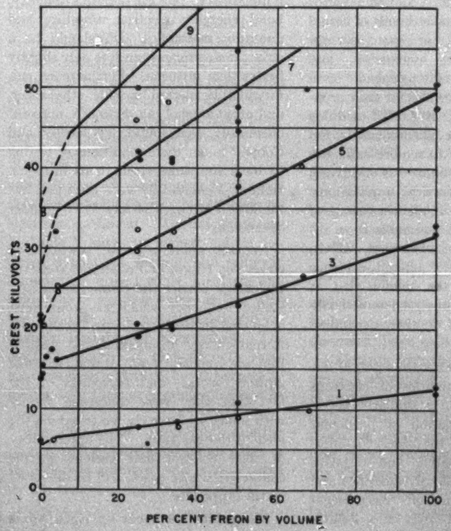
Toxicity as a Major Consideration

Fluorine compounds are of great interest in the extremes which they show. On one hand they are of considerable inertness and on the other there are some of great reactivity. A property which is not well understood is the toxicity of organic fluorine compounds. Some are among the least toxic of substances, while others are among the most poisonous known. The fully fluorinated fluorocarbons as a class are probably among the least toxic of organic compounds.[17]

Much work has been done in the Freons and on sulfur hexafluoride,[18] the latter stated definitely to be physiologically inert. Referring to Table I, the toxicities of the Freons is generally low as determined by accepted methods in the Underwriters Laboratories.

However, the relative toxicity after arcing is principally dependent on the formation of reaction products. For example, Freon-12 was said to produce free chlorine and fluorine under the action of corona.[19] Results[9] on arced SF_6 showed that definite low concentrations in air produce no serious physiological effect. For example, with 7.5 per cent or less of the arced gas, there was no harmful effect on rats; with higher concentrations up to 20 per cent and exposure for 10 minutes, slight signs of irritation in the lungs on autopsy after 24 hours were noted.

To our knowledge, work on the toxicity of other fluorocompounds, especially after electrical failure is incomplete and should be done before there is industrial



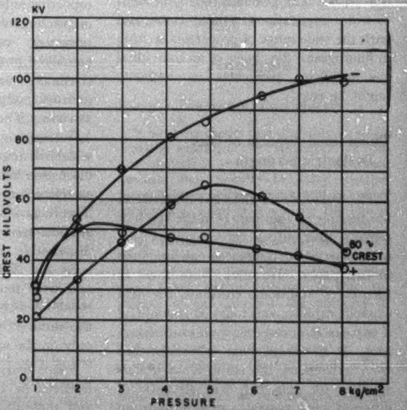


Figure 5 (left). 60-cycle sparking voltages in mixtures of Freon-12 and nitrogen versus Freon. Gap 0.05 inch, spherical electrodes. Pressure in atmospheres as noted on the curves. From reference 7

Figure 6 (above). 60-cycle and impulse breakdowns of 10 per cent Freon-12 and 90 per cent air versus pressure, kg/cm². Gap 10 millimeters, point-to-plane electrodes. From reference 15

application of the particular substance. It has been suggested that compounds containing only carbon and fluorine may produce less toxic products than SF$_6$.[11] Methods of purifying arced gases have been determined to be most practical and effective when absorption is used. Activated alumina as an absorbent was found efficient in removing the decomposition products of SF$_6$.[9,20]

General Outlook and Possible Applications

It appears that certain fluorogases may replace air in dry-type transformers. When a closed system is used, the advantages of greater insulation strength and cooling are obtained at little extra cost. Some fluorocarbons have already been used in high-voltage X ray apparatus and were later substituted by sulfur hexafluoride. This gas has also been used in medium voltage diagnostic apparatus in place of oil which had been used before. SF$_6$ has been used in experimental electrostatic generators, high-voltage cables, and capacitors.[21] An experimental transformer filled with SF$_6$ at 15 pounds gauge operated trouble-free for 2 years at 1.87 voltage over-load above rating.[9]

For indoor use, or in warm climates, the heavier fluorocarbons may prove advantageous after they have been more fully tested and evaluated. Continued testing of existing substances and new syntheses should be carried on of course. Specialized uses generally open the door to wider applications in many cases; and with the wide range of properties existing in fluorogases, the choice of an individual gas satisfying particular requirements might be easy.

Theoretical Aspects of Gas Dielectric Strength

The literature on dielectric phenomena has become so voluminous in the last half century that only a very small fraction of the pertinent references can be given in this paper. However, since these in turn carry additional citations, the field expands rapidly, and today it is doubtful if any one man can follow all of its divergent lines. Regarding whatever has been written on the subject, an understanding of the dielectric strength of different gases seems to remain upon an empirical basis.

Therefore, attention will be called to possible factors entering in the phenomenon of dielectric strength in a very broad way, fully realizing that some areas will thereby be neglected, but even worse, depth of presentation through detail will be sacrificed. However, there appears to be some need to clarify or define certain terms and to point out that dielectric strength is complex and always dependent on more than one property.

It would be well to find an explanation for the existing differences in the dielectric strength of various gases. Purely mechanistic theories dealing with the mean free path of the electron, limited by molecular cross sections and inelastic collisions, become inadequate when different gases are found, for example N$_2$O and CO$_2$ (both molecular weight of 44), which should have the same strength but do not.[22] The difference in acceleration of the electron by the applied voltage is therefore only one of the factors in dielectric failure.

The cumulative ionization theory of Townsend has beem employed for more than 40 years to account for differences in gases; a weak dielectric gas presumably produces many ions and an exponentially increasing number of electrons, through successive collisions, while a strong gas does not produce so many. The measure of the degree of ionization α has so seldom been reported in actual numbers (usually only as the ratio α/p) that to assess its direct bearing on dielectric strength is difficult. (Necessary primary data in the literature has been in many cases regrettably withheld. Confusion regarding numerical values of α arises through the pernicious habit of reporting only α/p, E/p, and pd as ratios or products without the details of actual pressures, voltages, or gap spacings employed.) Corona, avalanches, and streamers undoubtedly depend upon ionization of gas molecules for their propagation. The work of Loeb[23] and his co-workers on these phenomena is too widely disseminated to require extensive discussion here. Cobine[24] has effectively correlated many practical applications. The inclusion of α in all discussions of gas dielectric failure has made it an indispensable number. However, Hochberg and Sandberg[25,26] offered evidence and were flatly of the opinion that at breakdown the α of one gas is practically the same as that of another, depending only on the applied field.

Space charge, which is the result of differences in the mobilities of electrons and molecular ions, is similarly necessary in the explanation of particular phenomena such as the observed differences between positive and negative point corona as previously described and the effect of pressure on "electronegative" gases. In the latter case, Pollock and Cooper[3] asserted for SF$_6$ and CCl$_2$F$_2$ that an accelerated rate of positive ion formation with increasing pressure overcomes the stabilizing effect of relatively fewer negative ions (when the point electrode is positive), and breakdown can occur from a single streamer because of enhanced distortion of the field. They also indicate that factors such as the velocity of streamer propagation, streamer diameter, photoionization, and electron capture (all probably dependent on the specific gas) are also important considerations.

The chemistry of dielectric gases in relation to dielectric strength has been a matter for speculation and hypothesis since 1889. The almost invariable rise in dielectric strength when a given hydrocarbon, for instance methane CH$_4$, is converted to a halogen methane, such as CCl$_4$, has long been noted. The Freons are of course examples of this class. Other fluorocarbons, such as those of Wilson, Simons, and Brice[10] and Bashara[11] are members of homologous series, and the chemical similarity of hexafluoroethane (C$_2$F$_6$), octafluoropropane (C$_3$F$_8$), and decafluorobutane (C$_4$F$_{10}$) is so great that a simple linearity between the dielectric strength and the increasing density of these gases was found. It is only when gases are more distinctly different chemically that an accounting of relative dielectric strengths becomes more difficult. The case of CF$_4$ when compared with other stronger Freons such as CCl$_2$F$_2$ is illustrative. On the basis of symmetry, bond energies, electron affinities, and quantum mechanics, CF$_4$ should be a strong dielectric gas but it is only slightly better than nitrogen. Furthermore, sulfur chloride, methyl bromide, ethyl ether, and ethyl alcohol, all distinctly different chemically, were found by Charlton and Cooper[2] under their conditions of testing to have the same strength as nitrogen. Helium, a chemically very inert gas, has the lowest strength of any [3,12,13,22] under discussion.

Another point of view, traceable to either von Hipple or Frölich,[27,28] concerns the energy barrier and characteristic substance frequency which electrons must overcome in order to cause breakdown. Actual energy absorptions and dissipation are implied, if not stated, in these authors' works, and regardless of the fact that they were concerned mainly with solids, we have chosen to extend the idea to the gas state. The applicability of their concept and various specific characteristics of gases, will be discussed in the Appendix.

Meanwhile, comment on some terms

Table V. Some Values of the Atomic State of Selected Elements

	H	He	F	Cl	C	N	O	S
Relative electronegativity*	2.1		4.0	3.0	2.5	3.0	3.5	2.5
Electron affinity† $M+e\rightarrow M^-$	0.74	−0.53	4.1‡	3.8	1.5§	−0.6	2.8	1.5
Ionization potential‖ $M\rightarrow M^+ + e$	13.5	24.5	18.6	12.9	11.2	14.5	13.6	10.3
Atomic refractions¶	1.1	0.52	1.1	5.97	2.42	2.5	1.53	7.7
Atom periodic number	1	2	9	17	6	7	8	16
Atomic weight	1	4	19	35.5	12	14	16	32

* Pauling's original values.[29]

† From Massey,[32] values in electron volts.

‡ Two recent values are out of agreement: 3.18 ev (73 kilocalories per mole) from reference 44, and 3.57 ev (82.2 kilocalories per mole) from reference 45.

§ Massey's value, reference 32, page 18. Glockler & Sausville[46] found an experimental value of 2.1 ev with the heat of sublimation of carbon at 124 kcal.

‖ From various sources, in electron volts. See reference 47 or reference 48.

¶ Values used in organic chemistry. Nitrogen value in hydrazine N_2H_4, oxygen value in water H_2O, sulfur value in mercaptans RSH.

used to describe certain electrical properties of molecules is probably in order since the terminology of our subject is becoming more profuse and specialized.

Electronegativity, Electron Affinity, Ionization Potential, Appearance Potential, and Electron Trapping

"Electronegativity," the attraction of an atom for the electrons which bind it in a stable molecule,[29–31] has explained in part the high dielectric strength of some of the halocompounds. The halogen atoms have high electronegativities, see Table V, and high-strength gases produce negative ions on breakdown. In fact, any gas which does not form negative ions is unlikely to have exceptional dielectric strength, or conversely, all high-strength gases can produce negative ions. Whether this condition is causative or simply coincidental remains to be seen.

Electronegativity as the term is presently used, is rather a loose designation of an ability to consume free electrons.

"Electron affinity" is used more quantitatively to describe the process.[32]

$$M+e\rightarrow M^-$$

and indicates the energy change when an atom or molecule forms a negative ion by electron attachment. The halogen atoms in Table V have the highest values, followed by oxygen, carbon, and sulfur in diminishing order, while nitrogen and helium have negative values in line with the fact that their negative ions are extremely unstable or nonexistent. The electron affinities of high-strength gases have not been measured, but experimental methods for their determination could be used.

The "ionization potential," another property of a gas, should be an index of its dielectric strength, and is designated by the energy change in the process

$$M\rightarrow M^+ + e$$

Here, a positive ion and an electron are obtained from a neutral atom or molecule. Usually a highly energized electron is stated or assumed to be the initiator of the process and it is evident that electron avalanches can start in this way if enough molecules become involved in the action.

Practically all matter can produce positive ions and electrons, the potential of ionization varying from substance to substance. The suggestion of its value, however as an index of dielectric strength must be abandoned, at least as a single criterion since inspection (Table V) shows that the weakest dielectric gas, helium, has the highest ionization potential.

The "appearance potential" is a more recent term indicating the energy required to produce various, charged molecular fragments in the mass spectrometer. In this, a great many different ions are formed and identified. It is interesting to note[33] in passing that SF_6 going to SF_5^+ has the same appearance potential, 15.9 volts, as CF_4 going to CF_3^+ but SF_6 is the much stronger dielectric gas. These examples are different from that illustrating ionization potential, as previously defined, in that dissociation of the molecule preceded or accompanied the process, for instance

$$SF_6\rightarrow SF_5^+ + F^- \text{ (dissociation)}$$

and this required the concept of a dissociation energy or bond energy to be discussed later.

"Electron trapping" (used only in the sense of consuming; not as in the theory of metals or ionic crystals) is closely associated with electron affinity. In the preceding equation the energized electron (not shown) causing the dissociation of the SF_6 molecule rebounded with a loss energy (that used up in the SF_6 fragmentation).[8] An alternative occurrence is the inelastic collision of the electron with attachment to the molecule, and the same process as was shown for electron affinity applies. The conditions controlling electron trapping or attachment are:

1. The pressure of the gas.
2. The velocity of the colliding electron.
3. The molecular cross section of the particular gas molecules.
4. The electron affinity.
5. Probably the temperature.

In general, only slow electrons, that is, those with energies well under 5 electron volts (ev), will attach to molecules; faster electrons may attach but can cause extensive simultaneous dissociation. Since the distribution of velocities in any electron population affecting a gas probably includes many slow ones, it is evident that to have these consumed in forming even slower heavy negative ions should raise the voltage requirement for breakdown.

In all of the foregoing discussion, a few of the factors probably related in some way with dielectric strength have been very briefly noted or defined. The work of quantitative correlation, including empiricism and induction, has been privately in progress and if successful, may be reported at some future time. Additional theoretical bearings on the nature and properties of gases will be found in the Appendix. We must conclude that no single factor can be entirely responsible for dielectric strength; the combination of physical and electrical properties in a substance without doubt produces the resultant strength or weakness.

Appendix I

Further Theoretical Considerations

It has been shown that probably no single criterion exists for the explanation of relative dielectric strength. Isolated terms, however, may possibly be pulled together by applying a unifying concept of an energy summation, more or less along the line of reasoning of Fröhlich or von Hipple.[27,28]

A volume of gas is a rather complex system of forces. These can be considered as opposing the force of the applied electrical field. But in addition to this, the volume of gas also has the property of energy absorption, and by absorbing and dissipating energy, in one way or another, a higher potential must be applied to cause electrical failure. (An excellent monograph[16,34] considering all the physical properties of gases, dielectric strength omitted, has been recently published.)

A list of the strengthening forces in a volume of gas should include:

1. The cohesive forces among the molecules.
2. The cohesive forces among the atoms of the individual molecules.

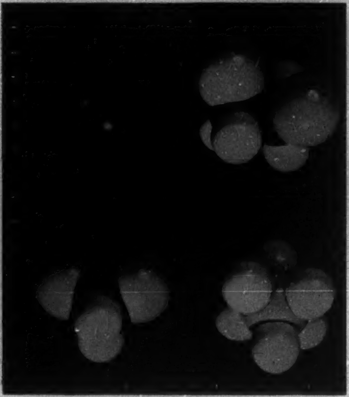


Figure 7. Molecular models of four gases. Divisions in Angstrom units

Top left, nitrogen; top right, Freon 114; bottom left, Freon 12; bottom right, SF_6

3. The cohesive forces of the electrons in the atoms.

Absorptions of energy can occur in a volume of gas through:

A. Thermal, translational, and rotational movement of the molecules, all kinetic
B. Atomic vibrations or oscillations of the atoms in the individual molecules.
C. Electron vibrations and excitation in the atoms, including structural resonance.
D. Induced, and permanent, dipole reactions to the applied field.
E. Electron attachment.
F. Chemical reaction after electron attachment or other ionization process.
G. Compression or electrostriction.

On the other hand a volume of gas also contains weakening forces and these are principally the repulsions existing among the molecules, atoms, and electrons. These should rightly be subtracted from the positive forces of cohesion. (Both forces are functions of the distance between particles, but in different degree.) Atomic and molecular inertia, if they diminish vibrational, rotational, and translational responses, should also be subtracted.

Most of the preceding general statements can be correlated with the present gases under consideration. Figure 7 shows four typical molecules illustrating differences in size, structural shapes, and chemical complexity of different dielectric molecules.

The van der Waal's equation of state provides an approximate measure of the cohesive forces among the molecules of a gas. The third column in Table VI lists the values obtained by using the original relationships of van der Waal on available critical data. A rough proportionality with dielectric strength is noted, excepting CHF_3 whose strength is out of agreement with an a value of 5.

Bond energies,[35] dissociation energies,[35] and stretching force constants [36,37] provide measures of the cohesive forces between atoms in a molecule. Available values are shown in Table VI columns 4, 5, and 6. That the values are for the individual bonds, as shown, is emphasized; they are only part of the total energy of the molecules.

The study of the electronic cohesive forces in a molecule is one of the problems of quantum mechanics. For our purpose, the ionization potential, which is the potential needed to remove an electron from the system, is a measure of the net cohesive force. (In this respect, the appearance potential of the "parent" ion in the mass spectrograph corresponds to the ionization potential.) Values for this are also listed in Table VI. The fact that the ionization potential by itself cannot be a true indication of dielectric strength has already been mentioned.

The aforementioned positive forces of cohesion are supplemented by positive absorptions of energy in the volume of gas. It is very likely that in the stress of an applied electrical field every frequency in the energy spectrum from the far ultraviolet to the far infrared can be present in some degree, depending on the manner and kind of voltage application. The total energy which a molecule can absorb in the energy field is always considered to be the sum of items A, B, and C, previously listed independently, plus interactions of A, B, and C, which are of lesser magnitude. The complete [34] expression for the available energy per molecule of an ideal gas with polyatomic linear molecules is

$$G = E_0 - (7/2)RT \ln T + RT \ln p + RT\Sigma \ln \times (1 - \epsilon^{-h\nu/kT}) - RT \ln (2\pi m)^{3/2}k^{5/2}) \times (h^{-3})\cdot(8\pi^2 I_k)(h^{-2}s^{-1})\pi g_e(g_n)$$

and for nonlinear molecule, such as Freon-12, the last term becomes

$$(8\pi^2(2\pi k)^{3/2}(I_A I_B I_C)^{1/2})(h^{-3}s^{-1})g_e\pi(g_n)$$

and this causes considerable trouble when

Table VI. Some Additional Properties of Various Gases

Formula	Relative Dielectric Strength	a van* der Waals	Bond† Energies, e.v.	Dissociation Energies, e.v.	Force‖ Constants, Dynes per Centimeter	Ionization** Potential, e.v.	Molecular** Refraction
H_2	<1	0.24	4.52	4.4776‡	5.76×10^5	15.4	2.05
He	≪1	0.034				24.5	.52
N_2	1.00	1.39	1.87	9.764‡	22.5×10^5	15.5	4.39
O_2	1±0.2	1.38	2.26	5.084‡	11.8×10^5	12.2	3.96
CHF_3	0.75	5.0	C.F 2.7 C.H 3.6	 C.H ca 3.9§	C.H 5.032×10^5 C.F 6.2×10^5		6.8††
$CHClF_2$	1.3	3.9	C.Cl 2.6	C.Cl 3.§	C.Cl 3.4×10^5		15.1
CF_4	1.1	1.2	C.Cl 2.6		C.F 6.246	17.8	10.1
$CClF_3$	1.4	3.6	C.Cl 2.6		C.Cl 3.458		11.7††
CCl_2F_2	2.4	5.4	C.Cl 2.6			11.7	20.1
C_2F_6	1.8	1.7	C.C 2.7		C.C ca 4.5		11.4††
C_3F_8	2.0	2.7	C.C 2.7				16.0††
C_4F_{10}	2.5	2.9	C.C 2.7				20.6††
SF_6	2–2.5	4.1	S.F 2.5		S.F 5.11¶	19.3	15.7
$CClF{=}CF_2$		10.5	C=C 3.9		C=C 9.8		15.8††
$CClF_2.CF_3$	2.8	6.5	probably same as other				16.3††
$CClF_2.CClF_2$	2.8	8.3	Freons				21.2††

* In (liter)^2 atmosphere, $a = 3p_c v_c^2 = 27R^2T_c^2/64p_c$

† Reference 35.

‡ Reference 49.

§ Reference 50.

‖ References 36 and 37.

¶ Calculated by reference 30 with $d = 1.58A$.

** Reference 41.

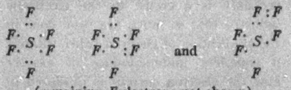
†† Calculated from the atomic refractions in Table V.

some of the quantities are not known. However, we may continue to make qualitative observations.

Thermal and translational movement of the molecules absorb energy if there is a source of heat and in actual numerical value probably is less than 0.005 electron volts per molecule. The rotational movement of a molecule is of a slightly higher order but still is probably less than 0.01 ev per molecule per degree of freedom.

Vibrational atomic oscillations absorb energy in the infrared and account for less than 0.1 ev per molecule per degree of freedom, that is, for a single vibration. If n is the number of atoms, the number of possible vibrations, that is degrees of freedom including the rotational, is given by $3n-6$ (or $3n-5$ if the molecule forms a straight line); and of these, $n-1$ are of the stretching type. The force constants of this type of vibration have been listed in Table VI and are related to the specific frequency of the vibration and to the masses of the connected atoms. In general, the greater the force constant, the greater energy absorption should be expected.

The electrons within the atoms are potentially the largest energy absorbers, ranging from about 1 ev upward to the ionization potential of the molecule. Energy calculations on the molecular electrons are difficult and uncertain except from spectrographic data. It is known that like the rotational and vibrational energies, the absorption of energy by electrons occurs in quanta. In a polyatomic molecule such as Freons, excited levels can be raised, if not directly by the external field, then by collisions. The acquired energy can be dissipated by radiation or by the vibrational excitation of the atoms which in turn can radiate at some lower frequency. This is different from another property possessed by some molecules called resonance[29] which is an electronic condition of oscillation between two or more atoms in a molecule, and is best thought of as a delocalization of particular electrons, so that several alternative electron bonds among the atoms can be visualized. Since the condition of resonance is a dynamic one, there is only a probability that the arrangement will take one particular form at any instant. The important effect of this condition is an increase in the total energy of the molecule; the molecule is more stable as a result of resonance. In SF_6 there appears to be a probability of resonance because the electron arrangements



(remaining F electrons not shown)

in terms of two covalent bonds and four semipolar singlets[38] are all possible, and it has been shown that there are no unshared electrons in the valence shells.[39] The fact that SF_6 has no permanent dipole moment does not exclude the possibility that some of the valence electrons may be in a state of resonance. The SF_6 molecule is generally agreed to be a regular octahedron in space with the sulfur at the center. The physical opportunity for resonance is therefore good. Finally, the strongest evidence is the arithmetic difference between the heat of formation,[40] 262 kilocalories per gram mole, and that obtained from Pauling's bond[38] energies, $6 \times 58.7 = 349$ kilocalories per mole. The difference, neglecting relatively small heats of fusion and vaporization, can be stated to be the resonance energy of stabilization, 87 kilocalories per mole or 3.78 ev. The Freons and fully saturated fluorocarbons have no such variation ability in their electronic arrangement.

Dipole Reactions

A molecule in an applied field becomes deformed or polarized, and this is a common property of all substances in varying degree from helium to very complex materials. Further, a certain part of the polarization is independent of the state of aggregation, that is whether the molecules are as a gas, liquid, or solid. From our point of view, the degree of polarizability is a measure of another means of energy absorption. In 1942 Hochberg and Sandberg[26] pointed out that the dielectric strength in a series of gases, including a few of the Freons, was proportional to the molecular refraction, which in turn is related to the polarizability through the Clausius-Mosotti-Debye equation. In Table VI the listed molecular refractions contain the electron refractions plus the atom refractions as given in Landolt-Börnstein tables,[41] and the proportionality with dielectric strength is evident. Helium is again found in its usual position as the weakest dielectric gas and a significant difference exists between CF_4 and CCl_2F_2. The effect of a permanent dipole, if present in a molecule, is an energy additional to the refractions shown; there is also dipole interaction which has a subtractive effect. These cannot be developed in discussion at this time; however, some known dipoles in Debyes are: $CHClF_2$, 1.4; $CClF_3$, 0.45; CCl_2F_2, 0.51; CF_4, 0.0; SF_6, 0.0 and comparisons in relative dielectric strength are interesting in this series.

"Electron attachment," since the pioneering work of Bradbury[42] on a variety of gases, has become of fundamental interest. The energy of the electrons in this process is wholly absorbed by the molecule and this fact should perhaps be emphasized. What the molecule, now an ion, does with the added energy depends on the particular molecule, and most of the energy will be dissipated in various vibrations previously described and by collisions with other particles or the wall of the vessel. Dissociation, or splitting of the molecule, however, can expend relatively more, and it is likely that all the fluorocompounds spontaneously dissociate on attaching an electron of sufficient energy as well as becoming simply negative with enhanced vibrations. Academically, a distinction can be drawn between simultaneous attachment and dissociation, and the possibility of these going on stepwise in finite time. It would seem that the stepwise process should be more acceptable on the basis of a common progenitor, the "parent" ion, even if this is very difficult to find after the reaction. For example, if Freon-14 forms the parent ion CF_4^- by attaching an electron, given sufficient energies, the following can then occur

$$CF_4^- \text{——} CF_3^\circ + F^-$$
$$CF_4^- \text{——} CF_3^+ + F^- + e$$
$$CF_4^- \text{——} CF_3^\circ + F^\circ + e$$

Of these, the first two are more likely than the last which is a higher energy reaction. These possibilities are only a few showing end products of an electron attachment. These products in turn can then enter into further electron reactions and by continuing to attach electrons the effect can be one of continuing energy absorption. Reaction possibilities of SF_6 have been briefly presented in a previous publication.[9]

The chemical, physical, and energy consideration of gases has not been exhaustively treated; much of the foregoing discussion has perhaps been presented uncritically. The fact that the problem of relative dielectric strength can be complicated has, however, been emphasized. The correlation of various factors, some applying generally and some under special conditions, is a real and challenging problem which may demand considerable time and effort.

References

1. Symposium on Fluorine Chemistry. *Industrial and Engineering Chemistry* (Easton, Pa.), volume 39, March 1947, pages 235–434.

2. The Dielectric Strength of Insulating Fluids, E. E. Charlton, F. S. Cooper. *General Electric Review* (Schenectady, N. Y.), volume 40, 1937, pages 438–42.

3. The Effect of Pressure on the Positive Point-to-Plane Discharge in N_2, O_2, CO_2, SO_2, SF_6, CCl_2F_2, A, He and H_2, H. C. Pollock, F. S. Cooper. *Physical Review* (New York, N. Y.), volume 56, 1939, pages 170–75.

4. Breakdown Studies in Compressed Gases, Alvin H. Howell. *AIEE Transactions* (*Electrical Engineering*), volume 58, May 1939, pages 193–206.

5. High-Pressure Gas as a Dielectric, Gordon C. Nonken. *AIEE Transactions* (*Electrical Engineering*), volume 60, December 1941, pages 1017–20.

6. D-C Breakdown Strength of Air and Freon in a Uniform Field at High Pressure, J. G. Trump, F. J. Safford, R. W. Cloud. *AIEE Transactions* (*Electrical Engineering*), volume 60, March 1941, pages 132–35.

7. The Electrical Strength of Nitrogen and Freon Under Pressure, H. H. Skilling, W. C. Brenner. *AIEE Transactions* (*Electrical Engineering*), volume 61, April 1942, pages 191–95.

8. Gaseous Insulation for High-Voltage Apparatus, G. Camilli, J. J. Chapman. *AIEE Transactions*, volume 66, 1947, pages 1463–70.

9. Gaseous Insulation for High-Voltage Transformers, G. Camilli, G. G. Gordon, R. E. Plump. *AIEE Transactions*, volume 71, part III, 1952, pages 348–57.

10. The Dielectric Strength of Gaseous Fluorocarbons, W. A. Wilson, J. H. Simons, T. J. Brice. *Journal of Applied Physics* (New York, N. Y.), volume 21, 1950, pages 203–05.

11. Some Fluorochemicals for Electrical Applications, N. M. Bashara. Proceedings of Symposium on Progress on Quality Electronic Components, Radio Television Manufacturers Association (Washington, D. C.), 1952, pages 30–36.

12. Spark Breakdown Potentials as a Function of the Product of the Pressure by the Plate Separation in A, N_2, and H_2 for Pt and Na Cathodes, Florence Ehrenkrans. *Physical Review* (New York, N. Y.), volume 55, 1939, pages 219–27.

13. The Role of the Cathode Surface in Sparking Phenomena, Harold Jacobs, A. P. LaRocque. *Physical Review* (New York, N. Y.), volume 74, 1948, pages 163–65.

14. Influence of Electrodes on D-C Breakdown in Gases at High Pressures, J. G. Trump, R. W. Cloud, J. G. Mann, E. P. Hanson. *Electrical Engineering*, volume 69, 1950, pages 961–64.

15. Gas Under Pressure as an Insulant,

R. E. Gossens. *Paper Number 117*, International Conference on Large Electric High-Tension Systems (Paris, France), 1948.

16. An Advanced Treatise on Physical Chemistry, Volume I (book), **J. R. Partington.** Longmans Green and Company, New York, N. Y., 1949.

17. Scientific and Utilitarian Value of Fluorine Chemistry, **J. H. Simons.** *Industrial and Engineering Chemistry* (Easton, Pa.), volume 39, 1947, pages 238–41.

18. The Toxicity of Sulfur Hexafluoride, **David Lester, Leon A. Greenberg.** Archives of Industrial Hygiene and Occupational Medicine (Chicago, Ill.), volume 2, number 3, 1950, pages 350–52. Compare page 337 for Freon-12.

19. A New Million-Volt X-Ray Outfit, **E. E. Charlton, W. R. Westendorp, L. E. Dempster, G. Hotaling.** *Journal of Applied Physics* (New York, N. Y.), volume 10, 1939, pages 374–85.

20. Effect of High Voltage Electrical Discharges on Sulfur Hexafluoride, **W. C. Schumb, J. C. Trump, G. L. Priest.** *Industrial and Engineering Chemistry* (Easton, Pa.), volume 41, 1949, pages 1348–51.

21. Gaseous Electric Insulation, **B. M. Hochberg.** *Elektrichestvo* (Moscow, USSR), volume 3, 1947, pages 15–19. Compare **Hochberg, Reinov,** *Doklady Akademii Nauk* (Moscow, USSR), volume 70, 1950, pages 837–8.

22. The Electric Strength of Gases Measured by Corona Discharge, **W. M. Thornton.** *Philosophical Magazine* (London, England), volume 28, 1939, pages 666–78. Compare **A. A. Bolsterli,** *Bulletin Schweizerischer Elektrotechnischer Verein* (Zurich, Switzerland), volume 22, number 11, 1931, page 247 for comparison of He, N_2 and CO_2, pressure up to 20 kilograms per square centimeter. Also **Y. Miyamoto,** *Archiv fur Electrotechnik* (Berlin, Germany), volume 31, 1937, pages 383–8.

23. Fundamental Processes of Electrical Discharge in Gases (book), **Leonard B. Loeb.** John Wiley and Sons, Inc., New York, N. Y., 1939. The Mechanism of the Electric Spark (book), **Loeb and Meek.** Stanford University Press, Palo Alto, Calif., 1941. Compare **Loeb, C. G. Miller.** *Journal of Applied Physics* (New York, N. Y.), volume 22, 1951, pages 614–21 for a corona study in pure N_2, O_2 and mixtures.

24. Gaseous Conductors (book), **J. D. Cobine.** McGraw-Hill Book Company, Inc., New York, N. Y., 1941.

25. Ionization of Gases and Their Breakdown Strength (in English), **B. M. Hochberg, E. J. Sandberg.** *Comptes Rendus de l'Academie des Sciences de l'URSS* (Moscow, USSR), volume 53, 1946, pages 511–14.

26. Compare **Hochberg, Sandberg.** *Journal of Technical Physics* (Moscow, USSR), volume 52, 1942, pages 65–75 for the pertinent statements on SF_6 and molecular refraction (in Russian).

27. Dielectric Breakdown in Ionic Crystals, **H. Frohlich.** *Physical Review* (New York, N. Y.), volume 56, 1939, pages 349–52.

28. Electric Breakdown of Solid Dielectrics, **A. von Hippel.** *Transactions*, Faraday Society (London, England), volume 42, supplement A, A General Discussion on Dielectrics, 1946, pages 78–90.

29. The Nature of the Chemical Bond (book), **Linus Pauling.** Cornell University Press, Ithaca, N. Y., 1940.

30. A Relation Between Bond Force Constants, Bond Orders, Bond Lengths, and the Electronegativities of the Bonded Atoms, **Walter Gordy.** *Journal of Chemical Physics* (New York, N. Y.), volume 14, 1946, pages 305–20.

31. A New Electroaffinity Scale, Together with Data on Valence States, **Robert S. Mulliken.** *Journal of Chemical Physics* (New York, N. Y.), volume 2, 1934, pages 782–93.

32. Negative Ions (book), **H. S. W. Massey.** Cambridge University Press, Cambridge, England, 1950.

33. Dissociation of SF_6, CF_4 and SiF_4 by Electron Impact, **V. H. Dibeler, F. L. Mohler.** *Journal of Research*, National Bureau of Standards (Washington, D. C.), volume 40, 1948, pages 25–9. Compare Ionization and Dissociation by Electron Impact in CCl_2F_2 and CCl_4 Vapor, **R. F. Baker, J. T. Tate.** *Physical Review* (New York, N. Y.), volume 53, 1938, page 684.

34. An Advanced Treatise on Physical Chemistry, *op. cit.*, pages 345–52.

35. A Revision of Some Bond-Energy Values, **H. A. Skinner.** *Transactions*, Faraday Society (London, England), volume 41, 1945, pages 645–62. Compare Symposium on Bond Strengths. *Journal of Physical Chemistry* (Baltimore, Md.), volume 55, 1952, pages 289–372.

36. Force Constants, **J. W. Linnett.** *Quarterly Review* (London, England), volume 1, 1947, pages 73–90.

37. Substituted Methanes VI Force Constants and Calculated Thermodynamic Properties of Some Trifluoromethanes, **C. E. Decker, A. G. Meister, F. F. Cleveland.** *Journal of Chemical Physics* (New York, N. Y.), volume 19, 1951, pages 784–88.

38. The Electronic Structure of Hexafluorides, **T. G. Pearson, P. L. Robinson.** *Journal of the Chemical Society* (London, England), 1933, pages 1427–28.

39. Hexacovalent Bond Orbitals, **George H. Duffey.** *Journal of Chemical Physics* (New York, N. Y.), volume 18, 1950, pages 128–29.

40. Thermochemistry of Chemical Substances (book), **F. R. Bichowsky, F. D. Rossini.** Reinhold Publishing Corporation, New York, N. Y., 1936.

41. Zahlenwerte und Funktionen (book). **Landolt-Bornstein.** Springer-Verlag, Berlin, Germany, volume 1, part 3, sixth edition, Molecules II, 1951, pages 509–17 for refractions; pages 359–68 for ionization potentials.

42. Formation of Negative Ions, Parts I and II, **Norris E. Bradbury.** *Journal of Chemical Physics* (New York, N. Y.), volume 2, 1934, pages 827–39.

43. Freon Compounds. *Technical Bulletin B-2*, Kinetic Chemicals Inc., Wilmington, Del.

44. On the Electron Affinity of Fluorine, **R. H. Johnston.** *Journal of Chemical Physics* (New York, N. Y.), volume 19, 1951, pages 1391–3.

45. Note on the Electron Affinity of Fluorine, **Bernstein, Metlay.** *Journal of Chemical Physics* (New York, N. Y.), volume 19, 1951, page 1612.

46. The Electron Affinity of the Carbon Atom, **Glockler, Sausville.** *Transactions*, Electrochemical Society (New York, N. Y.), volume 95, 1949, pages 282–91.

47. Atomic Energy States (book), **Bacher, Goudsmit.** McGraw-Hill Book Company, Inc., New York, N. Y., 1932.

48. Introduction to Chemical Physics (book), **J. C. Slater.** McGraw-Hill Book Company, Inc., New York, N. Y., 1939.

49. Dissociation Energies of Diatomic Molecules (book), **A. G. Gaydon.** John Wiley and Sons, Inc., New York, N. Y., 1947.

50. The Determination of Bond Dissociation Energies by Pyrolytic Methods, **M. Szwarc.** *Chemical Reviews* (Baltimore, Md.), volume 47, 1950, pages 75–173 or Bond Dissociation Energies, **M. Szwarc.** *Quarterly Review* (London, England), volume 5, 1951, pages 22 et seq.

No Discussion

The Nickel Base Indirectly Heated Oxide Cathode—A Review

A. M. BOUNDS
NONMEMBER AIEE

P. N. HAMBLETON
NONMEMBER AIEE

THE barium oxide indirectly heated oxide cathode came into wide use in 1927 and has since virtually dominated the field of ordinary receiving tubes and many other special types as well. The purpose of this article is to review its present status from the standpoint of the electronics engineer rather than that of the physicist. The role of the cathode alloy as well as some of the possibilities and shortcomings of such a cathode as used in the commercial production of electron tubes will be emphasized. In such a discussion it is necessary, first, to review the possible mechanisms of emission; second, to consider the performance characteristics of oxide cathodes; and third, to offer some comments upon the properties of the various cathode nickel alloys and their effect upon cathode performance.

The oxide cathode is perhaps the least-understood element in the modern electron tube, and indeed its functioning may be the least-understood phenomena in the entire science of electronics. It certainly is a credit to the electronic industry that so much has been done in the application of electron tubes without a complete understanding of exactly how the electron current is produced in the first place. Research work on oxide coated cathodes is a highly specialized and complex undertaking. The very large number of variables, the many steps necessary to construct experimental

---

Paper 53-174, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted February 8, 1952; made available for printing December 1, 1952.

A. M. Bounds and P. N. Hambleton are with the Superior Tube Company, Norristown, Pa.

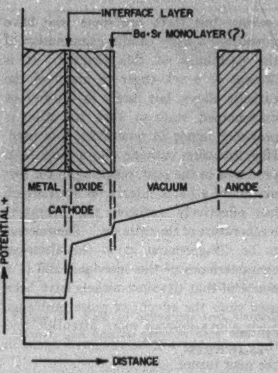


Figure 1. Schematic view through a diode tube with potential distribution approximation

tubes, and the difficulty in exact reproduction of these devices have restricted research to a very limited number of laboratories. Perhaps not more than a few dozen scientists in the entire world have directed intensive and long continued study to the oxide cathode. One of the earliest and most penetrating papers dealing with both the cathode base metal and its relation to the oxide coating was presented by Benjamin[1] in 1935. More detailed information has been developed by Sproull,[2] Eisenstein[3] and his associates in this country, and by Hermann and Wagner[4] in Germany, Loosjes and Vink[5,6] in Holland, Violet and Reithmuller[7] in France, and D. A. Wright[8] in England.

The oxide cathode has been spoken of as an "element" in the electron tube, and so it is usually considered by tube engineers. It really is a complex system composed of a metal cylinder or sleeve and a coating of alkaline-earth-metal carbonates initially held in place by a nitrocellulose binder. These carbonates of barium, strontium, and sometimes other additives must be carefully controlled as to shape, size, purity, and composition ratio. By a combination of evacuation and heating the carbonates are converted during tube processing to an adherent coating of oxide crystals, while the decomposed nitrocellulose binder and the carbon dioxide gases are withdrawn as they are generated. A carefully controlled schedule of high temperature heating, lower temperature aging, and actual cathode-current drain is required to complete activation. This process is designed to produce an optimum and stabilized electron emission from the cathode.

The precise nature of all the chemical and physical changes which occur on the coating, within the coating, and between the coating and the metal sleeve has long been the subject of speculation. An early theory held that activation was a result of electrolytic action within the mixed oxide coating, but it was not then understood that the so-called pure nickel base metal contained reducing agents which also entered into the reaction. Later, it was believed that a monolayer of free barium and strontium was formed on the outer surface of the oxide coating, and that barium impurity centers were distributed within the coating as well. Gradually the concept of the oxide coating acting as a semiconductor received considerable acceptance. More recently Loosjes and Vink[5,6] have advanced a dual-emission mechanism of pore conduction as an explanation of current passage through the coating resulting in liberation of free electrons (upon the application of a suitable positive external field). This theory is finding increasing support at the moment.

In any event, it is generally conceded that the low work-function of elemental barium and strontium in the oxide cathode provides for the release of a copious flow of electrons which are delivered to the anode when a suitable positive potential is applied. Figure 1 is a schematic view through a diode or 2-element tube showing an approximation of the potential drop when the cathode is emitting electrons to the anode.

For many reasons the oxide cathode is most widely used. It is a cheap and versatile mechanism, able to provide several hundred milliamperes of direct current for short periods, lower currents for as much as 100,000 hours, and high pulse currents (tens of amperes), depending on tube design, care, and control of manufacturing processes, coating materials and techniques and choice of metallic sleeve alloy. A comparison of typical emission current curves for oxide cathodes, tungsten, thoriated tungsten, and the "L" cathode taken from D. R. Hill[9] are shown in Figure 2.

Just as the oxide cathode is not a single element, but rather a whole system, so the characteristics of an indirectly heated oxide cathode may be separated into distinct categories, particularly with reference to the base metal alloy on which the cathode is built. These "performance characteristics" are interdependent to a degree but are sufficiently distinct to have real meaning. The following cathode performance characteristics are well worth careful attention

1. Rate of activation.
2. Emission levels, direct current and pulsed.
3. Rate of sublimation at a given operating temperature.
4. Tendency to form interface compounds of high electrical impedance.
5. Life characteristics.
   - Rate of free barium production and its influence upon grid emission.
   - Maintenance of primary emission as it contributes to maintenance of plate current and mutual conductance.
   - Maintenance of pulsed emission.
6. Heater-cathode leakage currents.

Coating and processing materials and techniques are extremely variable since each engineer in each tube plant has his own favorite materials and schedules. These are also dependent upon the equipment and tube types produced in the particular plant. Consequently, these phases, which are still a mixture of art and science, will be only touched upon, and the discussion will center on the more familiar nickel cathode alloys about which more is known with certainty.

How does the base metal affect these performance characteristics? The chemical reduction of earth-metal oxides by elements contained in the nickel sleeve has been mentioned. With one or two exceptions all cathode nickels contain controlled amounts of one or more of the elements magnesium, silicon, titanium, aluminum, carbon, and tungsten as chemical reducing agents for the alkaline earth-metal oxide coating. These may vary from as little as 0.01 per cent to as much as 4 per cent or more. When it is

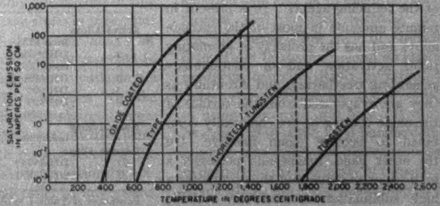


Figure 2. Saturation emission as a function of true temperature for L cathode and three other types

Table I. Commercial Cathode Alloys in Common Use, Significant Analysis Ranges

ASTM Grade	Producer Designation	C*	Mg	Si	Ti	Remarks
Active Alloys						
2	DH**-799	0.08 max	0.01/0.10	0.12/0.20	0.02 max	Si-Mg alloy
3	Inco §225	0.08 max	N.S.† (0.03)‡	0.15/0.25	N.S. (0.02)‡	Si-Mg alloy
4	DH-599	0.08 max	0.01 max	0.15/0.25	N.S. (0)‡	Si alloy
	Inco 330	0.08 max	N.S. (0.10)	0.10 max	N.S. (0.02)‡	Mg-Si alloy
Normal Alloy						
11	Inco 220	0.08 max	0.01/0.10	0.01/0.05	N.S. (0.02)‡	Mg-Si alloy
Passive Alloys						
21	DH-499††	0.05 max	0.01 max	0.01 max	0.01 max	Carbon deoxidized

*Maximum carbon in finished cathodes.
†Not specified.
‡Usual content.
§International Nickel Company, New York.
**Driver-Harris Company, Harrison, N. J.
††Patented.

considered that these alloys are melted in heats of from 700 to 20,000 pounds, the control of such small but important additives is evidently difficult and exacting. Each element and the various combinations of them affect the performance characteristics in different ways. While the effect of these elements may not be the only cause of the reduction of free barium, nevertheless the presence of various reducing agents in the nickel is easily demonstrated to have a profound effect on the performance characteristics of the cathode. Thus, the choice of proper cathode alloy for particular characteristics is more important than commonly understood.

The Reducing Elements

In such a short summary it is not necessary to review all of the major and minor effects of each of the chemical elements present in cathode nickel alloys, so only the most important and significant ones will be considered. Perhaps it should first be explained that there are a number of neutral and a few definitely poisoning elements, either unavoidably present or present for metallurgical reasons, which must be kept to reasonably low limits. The important ones are manganese, iron, copper, lead, boron, sulphur and phosphorus. These will not be considered here.

MAGNESIUM

Magnesium is the most common and important, as well as the most active of the reducing agents. It is present in minute amounts in most of the cathode alloys and has the property of promoting rapid activation and of continuously generating free barium and strontium as they are evaporated from the surface of the oxide coating or are chemically combined to become again inactive. On the other hand, magnesium possesses a high vapor pressure and thus promotes a high sublimation rate of the metallic sleeve often resulting in electrical leakage and shifting characteristics when the magnesium in the cathode alloy exceeds about 0.07 per cent. Magnesium does not produce a high impedance interface between cathode sleeve and coating, and therefore is satisfactory as an activator for pulsed cathodes.

SILICON

Silicon is next in commercial importance. Activation is not quite so rapid as with magnesium, but d-c emission life is long and the use of silicon has therefore become popular for a number of types of cathode alloys. Silicon may be present in quantities from 0.01 to 0.25 per cent. Unfortunately, silicon seems to promote the formation of an interfacial layer between cathode alloy and coating which has been identified as barium orthosilicate and which is characterized by high electrical impedance which, in turn, results in poor performance of tubes used for pulsed or cut-off purposes. Silicon displays a lower vapor pressure than magnesium and thus does not cause deposits on insulators resulting in electrical leakage.

TITANIUM

Titanium is used very sparingly and primarily for its metallurgical effects rather than for its reducing action on the coating, but there are some who have advocated a more careful examination of the usefulness of this element. It is present as a scavenger in many of the cathode alloys, but rarely in the free or uncombined state so that its effect is probably minor in quantities now used. High titanium cathode alloys have been produced in the past, resulting in the formation of a black interfacial layer which has effectively reduced the operating temperature of the cathode. There seems to be disagreement as to the electrical characteristics of this interface, and it is doubtful that titanium-nickels have been used since the advent of pulse problems. Such alloys deserve more attention and they probably will be reviewed again in the near future.

CARBON

Carbon apparently may act as a reducing agent on the coatings, particularly in the absence of any other highly reducing elements, but its role in cathode activation and life has been difficult to establish. Because of its high diffusion rate and formation of gaseous compounds, any effect by carbon is probably very short lived but may be of importance in activation of the cathode.

These elements have been traditional reducing agents for both activation and constant regeneration of barium and strontium during life. Now let us review the commercial and generally used cathode alloys (as shown in Table I and II). They are classified both on the basis of their activity and according to the particular reducing agent or agents which largely determine their performance characteristics. Table I shows those in common use, while Table II shows the entire list including many which are rarely used today.

The first action of Section "A" on Cathode Alloys of the American Society for Testing Materials (ASTM) Committee B4 was to divide the alloys, ac-

Table II. Commercial Cathode Alloys, Significant Analysis Ranges

ASTM Grade	Producer Designation	C*	Mg	Si	Ti	Remarks
Active Alloys						
1	Inco §224	0.08 max	0.01/0.10	0.12/0.20	0.02 max	Si-Mg alloy
2	DH**-799	0.08 max	0.01/0.10	0.12/0.20	0.02 max	Si-Mg alloy
3	Inco 225	0.08 max	N.S.† (0.03)‡	0.15/0.25	N.S. (0.02)‡	Si-Mg alloy
4	DH-599	0.08 max	0.01 max	0.15/0.25	N.S. (0)‡	Si alloy
5	DH-699	0.08 max	0.05/0.15	0.05/0.15	N.S. (0)‡	Mg-Si alloy
6	DH-399	0.08 max	0.01 max	0.15/0.25	N.S. (0)‡	Si alloy
	Inco 330	0.08 max	N.S. (0.10)	0.10 max	N.S. (0.02)‡	Mg-Si alloy
Normal Alloy						
11	Inco 220	0.08 max	0.01/0.10	0.01/0.05	N.S. (0.02)‡	Mg-Si alloy
Passive Alloys						
21	DH-499††	0.05 max	0.01 max	0.01 max	0.01 max	Carbon deoxidized
22	DH-999††	0.05 max	0.01 max	0.01 max	0.01 max	Special deoxidized

Footnotes of Table I apply to this table also.

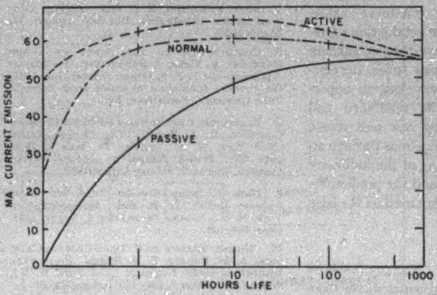


Figure 3. Rate of activation and emission levels typical of active, normal, and passive cathode base materials as tested in the standard ASTM diode

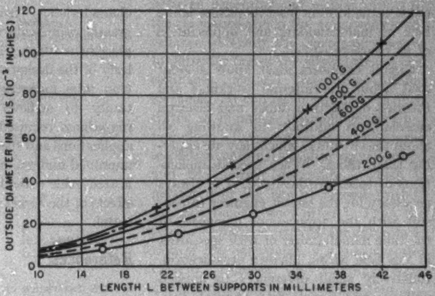


Figure 4. Relation of outside diameter and length between supports for round nickel tubing at several acceleration values and 775 degrees centigrade

cording to activity, into those which were "passive," "normal," and "active." This terminology is not exactly definitive when related to performance characteristics but does classify the alloys in a general way with respect to rapidity of activation and level of d-c emission during life. The rate of activation is of commercial importance to tube manufacturers but of little importance to the tube user.

The passive alloys show extremely slow activation, relatively low levels of emission, and extremely long life. The normal alloys show relatively rapid activation, higher levels of emission, and sometimes shorter life. The active alloys are rapid in activation, give slightly higher d-c emission and related tube characteristics, and when used under proper conditions give fairly long life. These general characteristics are shown in Figure 3. However, it is our belief that all but the passive alloys are probably outmoded by today's requirements which entail better control of sublimation characteristics, complete freedom from interface impedance, longer life, and better strength.

Newer alloys which have been recently introduced are ones containing aluminum as the activating agent in one series, tungsten as both a reducing element and a strengthening element in another series, and cobalt in still another. Now the use of these three elements is by no means new, since they are long familiar as ingredients in filamentary alloys. However, it was necessary to drastically alter the filamentary alloys, and in general to greatly reduce the content of these elements before the alloys would behave properly for indirectly heated cathodes. The new compositions are, therefore, quite distinct from the filamentary alloys.

The advantages of aluminum are evident from the characteristics of the element itself: activation is relatively rapid, d-c emission is high and long-lived, and sublimation has proved to be very low. The interfacial compounds, if present, are of low electrical impedance. In fact we may say here that in our experience only silicon has shown any effect in the creation of an interfacial layer of high impedance. Of course most of the reducing agents cause the formation of a visible interface layer, but these films do not necessarily show electrical impedance. There are indications that the absence of all reducing elements is necessary to the best pulsed performance, that is, passive cathode base metal is indicated.

In addition to doubling the hot strength of a cathode sleeve, the addition of approximately 4-per-cent tungsten seems to enhance the "reducing potential" of the cathode alloy. Tungsten-nickel alloys have shown reasonably rapid activation, long life under d-c emission conditions, relatively high emission, and complete freedom from interface impedance where the silicon content was sufficiently low. However, tungsten is a strategic element, scarce and high in price, so that tungsten-nickel alloys must be considered as premium ones for special purposes only.

Problems of reliability and shock and vibration resistance are specters which haunt tube engineers these days. Reliability is a problem not only of mechanical perfection of assembly but also one of long life with freedom from shifts in tube characteristics. The aluminum alloy for indirectly heated cathodes is expected to be a definite contribution to cathode life and stability of characteristics. On the other hand, ruggedization requires stronger cathode alloys which are, at the same time, suitable with respect to all of the cathode characteristics previously described; not a small order by any means.

Much improvement in shock resistance of cathodes can be effected with existing alloys by the application of simple and well-known laws governing the strength of shapes and materials. In considering this problem it must be understood that the heavier the cathode the less its resistance to deformation under shock loading. By the same laws it is almost elementary that the larger the diameter and the shorter the length between supports, the stronger is the cathode cylinder. And yet, these simple facts apparently are often ignored or forgotten in tube design work. In connection with this problem we have developed cathode strength curves which should enable a tube designer to predict with considerable accuracy the resistance of an indirectly heated cathode to shock loading. Figure 4 shows these curves for cathodes operating at a temperature of 775 degrees centigrade. The Bureau of Ships, Department of the Navy, has permitted publication of this small portion of recent contract development work by the Superior Tube Company in the hope that it will be an aid in the design of more rugged and reliable tubes.

In conclusion, it is desirable to emphasize a number of points which derive from the foregoing discussion. The complex oxide cathode structure is temperature dependent, not only as to electron emission but also as to chemical, metallurgical, and physical reactions. It is not an inexhaustible well of electrons; it must not be driven beyond its electron producing capacity simply by raising the temperature, raising the anode voltage, or using a positive grid bias. It cannot be expected to stand up if bombarded by positive ions which act like cannon balls in destroying the emissivity of the coating. Many of the apparent weaknesses of the oxide cathode stem from its misuse and a lack of understanding and appreciation of its characteristics.

Circuit engineers must know a great deal about the internal workings and theories of electron tubes and electron emission, in addition to knowing all about the circuits in which they are applying these tubes. When special applications of tubes are involved, it is always necessary for the circuit engineer to consult with the applications engineers of the tube manufacturer or with specialists in this field within his own organization. Relatively small differences in applied voltages and signals may be vital; they may change the tube life from 10,000 hours to a matter of hundreds of hours in a most unexpected fashion. To achieve reliability tubes should be underrated as much as is economically possible. Tubes must be processed for the application for which they are intended.

It must also be recognized that the complexity of materials and processes which are incorporated in the vacuum tube make it difficult both to substitute new alloys in old designs and to really evaluate the effects of these new materials. It is also well known that slight changes in tube characteristics, from whatever cause, greatly complicate the production of replacement tubes. Nevertheless, particularly in the design of new tubes, the electron tube industry now has an opportunity to achieve the reliability and ruggedness required by new and severe applications and specifications through an improved understanding of the indirectly heated oxide cathode and the predictable effects of the nickel alloy used as the base metal.

References

1. The Influence of Impurities in the Core Metal on Thermionic Emission from Oxide Coated Nickel, **M. Benjamin.** *Philosophical Magazine* (London, England), series 7, volume 20, number 131, July 1935, pages 1–24.

2. Investigation of Short Time Thermionic Emission from Oxide-Coated Cathodes, **L. Sproull.** *Physical Review* (New York, N. Y.), volume 67, 1945, pages 166–78.

3. Leaky Condenser Oxide Cathode Interface, **A. Eisenstein.** *Journal of Applied Physics* (New York, N. Y.), volume 22, number 2, February 1951, pages 138–48.

4. The Oxide-Coated Cathode, Volume 1 (book), **G. Hermann, S. Wagener.** Chapman and Hall, London, England.

5. Conduction Mechanism in Oxide-Coated Cathodes, **R. Loosjes, H. J. Vink.** *Phillips Research Reports* (Eindhoven, Holland), volume 4, number 6, 1949.

6. Conduction Processes in the Oxide-Coated Cathode, **R. Loosjes, H. J. Vink.** *Phillips Technical Review* (Eindhoven, Holland), volume 11, March 1950.

7. Contribution to the Study of Oxide Cathodes, **V. Violet, J. Reithmuller.** *Annales de Radioelectricité T. S. F.* (Paris, France), July 1949. (In French; translation available from Superior Tube Company, Norristown, Pa.)

8. Electrical Conductivity of Oxide Cathode Coatings, **D. A. Wright.** *The Physical Review* (New York, N. Y.), volume 78, May 1, 1950, page 355. *British Journal of Applied Physics* (London, England), volume 1, June 1950.

9. New Cathode Design Improves Tube Reality (book), **D. R. Hill.** *Electronics* (New York, N. Y.), volume 24, number 8, August 1951, pages 104–106.

10. Nickel Alloys for Oxide-Coated Cathodes, **A. M. Bounds, T. H. Briggs.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 39, number 7, July 1951, pages 788–99.

11. Application of Thermodynamics to Chemical Problems Involved in the Oxide Cathode, **A. H. White.** *Journal of Applied Physics* (New York, N. Y.), volume 20, September 1949, page 856.

12. The Life of Oxide Cathodes in Modern Receiving Valves, **G. H. Metson, S. Wagener, M. F. Holmes, M. R. Child.** *Proceedings*, Institution of Electrical Engineers (London, England), volume 39, part III, number 58, March 1952, pages 69–87.

13. Correlation of DC and Microsecond Pulsed Emission from Oxide-Coated Cathodes, **Frank A. Horak.** *Journal of Applied Physics* (New York, N. Y.), volume 23, number 3, March 1952, pages 346–50.

14. Activation of High Vacuum Oxide-Cathode Valves, **G. H. Metson.** *Vacuum* (London, England), volume 1, number 4, October 1951, pages 283–93.

No Discussion

Application of Electric and Acoustic Impedance Measuring Techniques to Problems in Diathermy

HERMAN P. SCHWAN
ASSOCIATE MEMBER AIEE

EDWIN L. CARSTENSEN
NONMEMBER AIEE

DIATHERMY, or heating of deep tissues, must be accomplished through the absorption of radiant energy in the region to be heated. Conduction methods such as infrared or hot packing fail because of the rapid convection of heat by the flow of blood. Conventionally diathermy has come to be synonymous with high-frequency electromagnetic heating. Recently, however, ultrasonic energy has been introduced to accomplish a somewhat similar clinical purpose.

The effectiveness of these forms of radiation may be evaluated largely in terms of these physical quantities: the depth of penetration, reflection coefficient, and beaming angle, all of which can be computed from the frequency dependent electrical and acoustic impedance of the biological medium.

Rather extensive measurements of the electric impedance of tissues and blood have been carried out in the region up to 100 megacycles by Osswald[1] and by Schaefer[2] and above 1,000 megacycles recently by England[3] and by Herrick and co-workers.[4] Actually, as will be shown, the most promising region for diathermy lies between these two extremes.

For measurements in the frequency range 100–1,000 megacycles a special resonance method[5] has been adopted, which is a modification of a technique first discussed by Chipman.[6] An open-wire transmission line, see Figure 1, is loaded on one side by the sample under investigation and terminated on the other end by a movable short-circuiting plate. The line may be excited either capacitatively or inductively at a point between sample and short-circuiting end. The modulated current is picked up by a small loop protruding from the movable short-circuiting plate. This current is subsequently detected, amplified, and metered. Figure 2 shows the variation of this current as a function of the position of the short circuit for the case where the reflection coefficient of the sample is approximately one half. The points were obtained experimentally. The solid line is theoretical. The slight discrepancy between experiment and theory away from resonance arises from direct pickup

---

Paper 53-137, recommended by the AIEE Electrical Techniques in Medicine and Biology Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted October 22, 1952; made available for printing December 29, 1952.

Herman P. Schwan and Edwin L. Carstensen are with the University of Pennsylvania, Philadelphia, Pa.

The authors wish to thank the National Foundation for Infantile Paralysis, Inc., for their grant.

between generator and pickup loop. In the resonance region the agreement is excellent. Actually the discrepancy serves to illustrate one of the advantages of this resonance method over conventional methods. It can be shown that, as illustrated in Figure 3, the voltage distribution along a transmission line as measured in the conventional standing wave method and the current distribution which is measured in the new resonance method are reciprocals of each other.[6] Thus, the resonance method requires the determination of the bandwidth of a maximum of current, whereas the standing wave method necessitates an investigation of a minimum of voltage.[7] Hence, the resonance method is less sensitive to stray fields and harmonic disturbances than the standing wave method. As a consequence, it is possible to measure with accuracy sample impedances which depart from the characteristic impedance of the line by factors up to 1,000.

By comparison with coaxial lines, open-wire transmission lines are flexible, accessible, and simple to construct. However, because of susceptibility to external disturbances and undesirable modes of excitation (as between system and ground), they are seldom used. The advantage of the resonance method, however, is that it makes it possible to obtain a high degree of accuracy with open systems. Measurements of the impedance of water have been conductance by the authors in the range from 36 to 200 centimeters, using an open-wire resonance system. The results are in good agreement with theoretical expectations.

Figure 4 shows measurements in this range, which have been obtained for blood, along with the results of other workers at higher and lower frequencies. The results are expressed in terms of dielectric constant and specific resistance. It can be seen that the data form a consistent pattern. The variation with frequency shown here is attributed to two different mechanisms. The existence of cell membranes in the blood causes a Maxwell-Wagner dispersion in the neighborhood of 1 megacycle.[8,9] This produces the change in the dielectric constant curve, which begins to appear, above 1 meter. The change below 30 centimeters results from Debye relaxation of the water molecules.[10] The dielectric constant between 30 and 100 centimeters is somewhat lower than that of water—54 as against 80 for water. This is explained by the presence of protein molecules.

Reflection and absorption coefficient α have been computed from these data and are presented in Table I. The absorp-

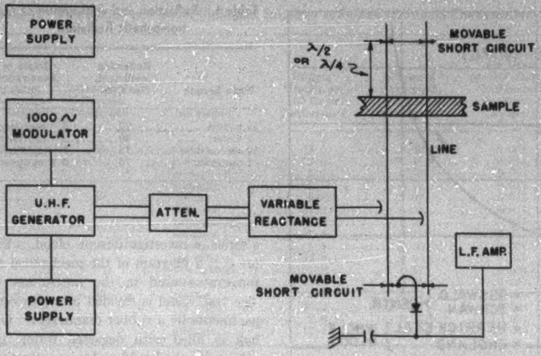


Figure 1. Resonance line for impedance measurements between 100 and 1,000 megacycles

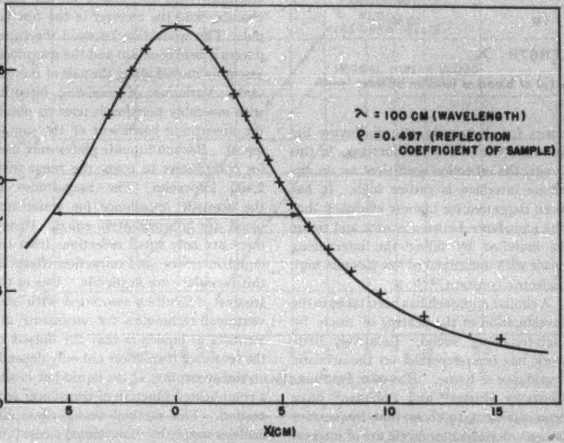


Figure 2. Resonance curve (points are experimental, curve calculated)

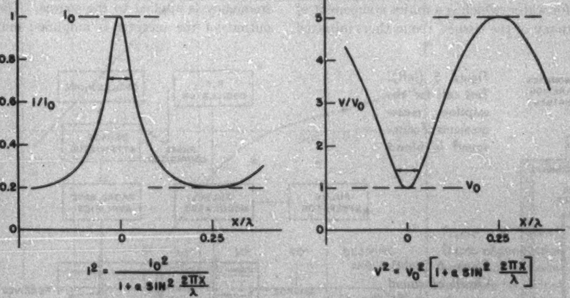


Figure 3. Comparison of resonance curve with standing wave pattern

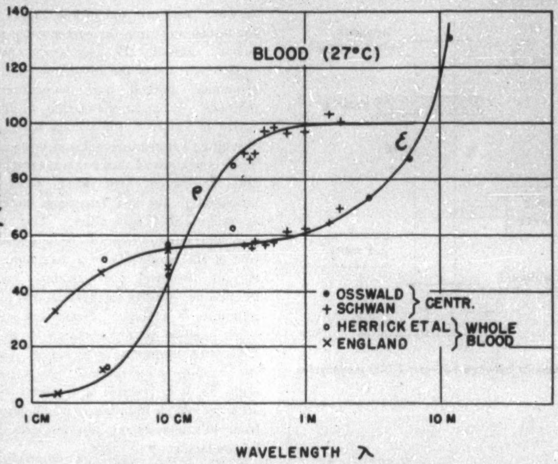


Figure 4. Dielectric constant (ε) and resistivity (ρ) of blood as function of wave length

Table I. Reflection and Absorption of Electromagnetic Radiation

Wave Length	Reflection Coefficient, Per Cent	Depth of Penetration (1/α)
10 meters	92	5 centimeters
1 meter	80	2 centimeters
10 centimeters	75	7 millimeters
1 centimeter	70	0.8 millimeter

tion is expressed in terms of depth of penetration of the intensity, that is, the depth in tissue at which intensity reaches 1/*e* of its surface value. Since cell membranes, proteins, and saline are from the electrical point of view the major constituents of all tissues of high water content, it can be said that the results for blood are characteristic of such tissue in general. The depth of penetration here decreases with decreasing wave length. For wave lengths less than 10 centimeters, the penetration is not significantly higher than could be obtained from surface heating. Use of wave lengths greater than 3 meters, on the other hand, is undesirable. If such relatively low-frequency energy is applied in form of electromagnetic radiation, any beaming is practically impossible. And, if it is supplied by high-frequency current, excessive selective heating of the subcutaneous fat layer results.[11] We therefore conclude that the optimum range for electromagnetic diathermy lies between 30 and 100 centimeters. In this range, the reflection coefficient for an air-tissue interface is rather high. It has been suggested for highest efficiency that the impedance between source and tissue be matched by filling the intervening space with a material of low loss and high dielectric constant.[12–15]

A similar approach has been taken in the investigation of the heating of tissue by high-frequency sound. Relatively little work has been reported on the acoustic impedance of tissue. However, measurements by Hueter[16] and Pohlman[17] have been sufficient to show that frequencies in the order of 1 megacycle are of interest for diathermy.

In an attempt to determine the cause of the relatively high absorption of sound observed for tissue, as compared with that for water, which is a major component of many of the tissues, the authors initiated a series of investigations on blood. Figure 5 is a diagram of the mechanical arrangement used in the measurements. The test vessel is divided into two compartments by a rubber diaphragm. One half is filled with degassed water, the other with the liquid under investigation. The transducers are mounted on a sliding assembly with the source in the water chamber and the receiver in the test liquid. The separation between the transducers is held constant and the transducer assembly moved along the axis of the test tank. Variation of receiving intensity with assembly position is used to obtain the absorption coefficient of the sample liquid. Barium titanate plates were used for transducers to cover the range 300–2,400 kilocycles. The magnitudes of the acoustic impedance for water and blood are approximately equal. Hence, there are only small reflections from the liquid interface, and refraction effects at this boundary are negligible. One of the greatest difficulties associated with conventional techniques for measuring absorption in liquids is that the output of the receiving transducer not only depends on the absorption of the liquid but is also a complicated function of transducer separation. This method avoids these difficulties simply by maintaining a constant transducer separation.

Effectively free field conditions are obtained through the use of pulsing techniques, see Figure 6. A pulse of radio frequency is applied to the source. The output of the receiver is amplified and

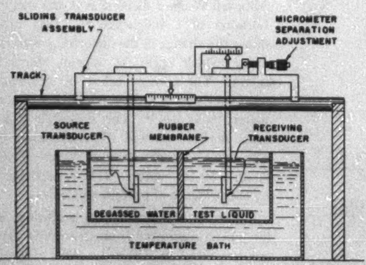


Figure 5 (left). Test cell for absorption measurements of ultrasound in blood

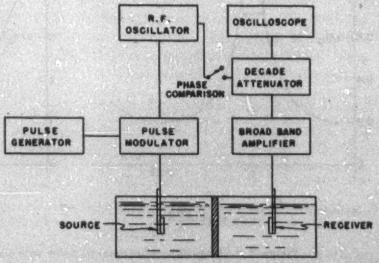


Figure 6 (right). Circuit of sound measuring equipment

presented on an oscilloscope. The relative output intensity is measured by use of a decade attenuator which can be adjusted for constant signal at the oscilloscope.

For velocity of sound determinations, the wave length in the test liquid is measured by comparison of phase of the radio-frequency output of the receiver with a direct signal from the generator while the transducer separation is varied by micrometer control.

Figure 7 shows absorption as measured for the red cell residues of centrifuged human blood and for plasma.[18] The absorption is very nearly a linear function of frequency. The velocity of sound in blood is independent of frequency and approximately equal to that of water. In seeking the source of the absorption of ultrasonic energy by blood, it should be noted that the absorption in plasma which contains no cells is of the order of 25 times that for water. Further investigation as summarized in Table II[19] shows that the absorption goes in direct proportion to the protein content for blood and its components. Here the absorption, as measured at various frequencies and temperatures for concentrated red cells, plasma, and solutions of hemoglobin and albumin, is expressed in terms of absorption per unit quantity of protein present in the solution. The agreement in the results for various solutions shows that the proteins, whether in cells or in solution, are responsible for the absorption. Further, the protein content of many of the solid tissues is high enough to account for a large part of the observed absorption on this simple basis.

The actual situation in the solid tissues is somewhat more complicated. There also, however, the absorption appears to be a linear function of frequency and runs only one order of magnitude higher than that of centrifuged blood. Depths of penetration for solid tissues are of the order of 4–8 centimeters at 1 megacycle. Hence, frequencies up to approximately 2 megacycles can be useful for diathermy.

According to clinical reports, high-frequency sound can be used for many of the same applications as electromagnetic diathermy. It will be interesting to compare the two forms of radiation quantitatively on a specific problem, namely, that of the localized heating of the deep tissues. The directional properties of a piston source of either electromagnetic or acoustic energy depend upon the ratio of wave length to piston radius. The wave length of sound in its useful range is considerably smaller than that of electromagnetic radiation and thus sound

Table II. Absorption of Various Biological Solutions as Related to Protein Concentrations

Frequency, Megacycles per Second	Temperature, Degrees Centigrade	Substance	Protein Grams, per 100 Cubic Centimeters	Measured Absorption, Decibels per Centimeter	Absorption per Gram per 100 Cubic Centimeters, Decibels per Centimeter
1.2	20	plasma	5.5	0.08	0.014
		red cells	32.0	0.40	0.013
		hemoglobin	7.0	0.09	0.013
		albumin	12.5	0.16	0.013
1.2	40	plasma	5.5	0.07	0.012
		red cells	32.0	0.34	0.011
		hemoglobin	7.0	0.08	0.011
		albumin	12.5	0.14	0.011
2.4	20	plasma	5.5	0.16	0.030
		red cells	32.0	0.94	0.029
		hemoglobin	7.0	0.17	0.024
		albumin	12.5	0.34	0.027
2.4	40	plasma	5.5	0.13	0.023
		red cells	32.0	0.77	0.024
		hemoglobin	7.0	0.15	0.021
		albumin	12.5	0.29	0.023

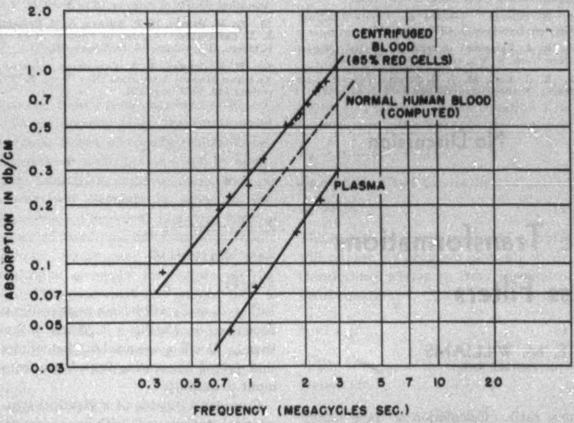


Figure 7 (above). Ultrasonic absorption of blood and plasma as function of frequency

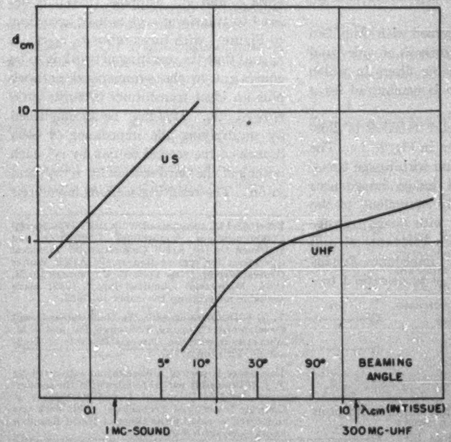


Figure 8 (left). Depth of penetration as function of beaming angle for various types of radiant energy may be more sharply beamed. This is shown graphically in Figure 8.[20] Here depths of penetration are plotted against wave length both for acoustic and electromagnetic radiation. The corresponding beam angle for a 5-centimeter radiator is also indicated along the abscissa. Over almost all of the useful range of sound the beam suffers practically no divergence. With even the shortest permissable wave lengths of electromagnetic radiation, however, the beaming is poor. From this analysis, it may be concluded that ultrasound is superior to electromagnetic radiation for the purposes of localized deep heating.

Summary

Measuring techniques are described for the dielectric constant and resistivity of biological material at ultrahigh frequencies and for the absorption of ultrasound. Electrical and acoustical data are presented for blood. The significance of these data with respect to the mechanism of absorption of the radiant energy is discussed. From a comparison using the electrical and acoustical impedance of tissue, it is possible to predict advantages of ultrasound over electromagnetic diathermy for the purpose of localized deep heating.

References

1. **K. Osswald.** *Hochfrequenztechnik und Elektroakustik* (Leipzig, Germany), volume 49, 1937, page 40.

2. **H. Schaefer.** *Zeitschrift fuer ges experimentelle Medizin* (Berlin, Germany), volume 100, 1937, page 706.

3. **T. S. England.** *Nature* (London, England), volume 166, 1950, page 480.

4. **J. F. Herrick.** *Proceedings*, American Physiological Society (New York, N. Y.), 1950.

5. **H. Schwan.** *Annalen der Physik* (Leipzig, Germany), volume 5, 1950, page 253.

6. **R. A. Chipman.** *Journal of Applied Physics* (New York, N. Y.), volume 10, 1939, page 27.

7. **K. S. Knol, M. J. O. Strutt.** *Physica* (The Hague, Netherlands), volume 9, 1942, page 577.

8. **H. Fricke.** *The Physical Review* (New York, N. Y.), volume 24, 1924, page 575.

9. **H. Fricke.** *The Physical Review* (New York, N. Y.), volume 26, 1925, page 678.

10. **G. Baez.** *Physikalische Zeitschrift* (Leipzig, Germany), volume 40, 1939, page 394.

11. **H. Schaefer.** *Zeitschrift fuer ges experimentelle Medizin* (Berlin, Germany), volume 98, 1936, page 257.

12. **B. Rajewsky, H. Osken, H. Schaefer.** *Naturwissenschaften* (Berlin, Germany), volume 25, 1936, page 520.

13. **A. Esau, J. Paetzold, E. Ahrens.** *Naturwissenschaften* (Berlin, Germany), volume 24, 1936, page 520.

14. **H. E. Hollmann.** *Hochfrequenztechnik und Elektroskustik* (Leipzig, Germany), volume 50, 1937, page 81.

15. **H. E. Hollmann.** *Strahlentherapie* (Berlin, Germany), volume 64, 1939, page 691.

16. **T. H. Hueter.** *Naturwissenschaften* (Berlin, Germany), volume 9, 1948, page 285.

17. **R. Pohlman.** *Physikalische Zeitschrift* (Leipzig, Germany), volume 40, 1939, page 159.

18. **E. Carstensen, K. Li, H. Schwan.** *Journal*, Acoustical Society of America (New York, N. Y.).

19. **G. M. Piersol, H. P. Schwan, R. B. Pennell, E. L. Carstensen.** *Archives of Physical Medicine* (Chicago, Ill.), volume 33, 1952, page 327.

20. **H. P. Schwan, E. L. Carstensen.** *Journal*, American Medical Association (New York, N. Y.), volume 149, 1952, page 121.

No Discussion

Maximum Impedance Transformations in Band-Pass Filters

T. J. O'DONNELL
NONMEMBER AIEE

E. M. WILLIAMS
MEMBER AIEE

Synopsis: Design of impedance transforming band-pass filters for use at radio frequencies often involves the criterion of maximum realizable impedance ratio with a minimum number of sections. A fundamental method is described for achieving this design criterion and illustrated for some simple cases.

WIDE-BAND impedance transforming elements for use at frequencies up to about 10 megacycles employ devices which are refinements of the ordinary iron core transformer. Outside this frequency range wide-band transformations commonly utilize impedance transforming wave filters. Design of such filters is derived from Norton's[1] equivalence between a dissymmetrical band-pass filter and a symmetrical section plus an ideal transformer. Various applications of Norton's equivalence have been developed by Gladwin,[2] Belevitch,[3] and Rowlands.[4] It is evident from these works that there are inherent limits to realizable band widths for any impedance ratio. Calculation of these limits, however, has been an indirect and involved procedure.

This paper is concerned with simplified direct methods for synthesis of wide-band impedance transforming filters in which the impedance ratio is maximized for a specified bandwidth.

Norton's equivalence relation is illustrated for a T-section in Figure 1. The dissymmetrical section with image transfer function θ and image impedances Z_{I_1} and Z_{I_2} is exactly equivalent to the symmetrical section with image transfer function (propagation constant) θ' and image (characteristic) impedance Z_I plus an ideal transformer of turns-ratio 1 to r provided that

$$\theta=\theta'$$

$$Z_{I_1}=Z_I$$

$$Z_{I_2}=r^2Z_I$$

It may be shown that these conditions hold if

$$Z_A=Z_1-Z_2(r-1)$$

$$Z_B=r^2Z_1+r(r-1)Z_2 \qquad (1)$$

$$Z_C=rZ_2$$

Either Z_A or Z_B will have a negative term depending on whether r is greater or less than 1. It will be assumed in what follows that $r>1$ (increasing impedance from input to output).

Consider a cascade of n identical symmetrical sections each with image parameters Z_I and θ'. Suppose that it is desired to transform each section according to Figure 1 with turns-ratios $r_1, r_2, r_3, \ldots r_n$, and that the resulting network is to be equivalent to the symmetrical network plus an ideal transformer of turns ratio $r_1 r_2 r_3 \ldots r_n$. This may be accomplished by multiplying the impedance of each branch of the second section by r_1^2, each branch of the third section by $r_1^2 r_2^2$, and so on. The resulting sections have their

---

Paper 53-143, recommended by the AIEE Special Communications Applications Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted July 7, 1952; made available for printing December 10, 1952.

T. J. O'DONNELL is with the Gulf Research and Development Company, Pittsburgh, Pa., and E. M. WILLIAMS is with the Carnegie Institute of Technology, Pittsburgh, Pa.

This paper is part of a dissertation submitted by T. J. O'Donnell in partial fulfillment of the requirements for the Degree of Doctor of Science at Carnegie Institute of Technology. This work was supported in part by the Office of Naval Research under Contract *NR-200(00)*.

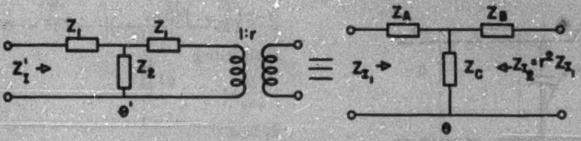


Figure 1. Norton's equivalence between a symmetrical T-section transformer combination and a dissymmetrical T-section

image impedances multiplied by the corresponding multiplier. Their image transfer functions remain unchanged. They may then be transformed by use of equation 1 and the transformed sections may be cascaded on the "image basis." The cascaded network will have image parameters Z_{I_1}, Z_{I_2}, and θ given by

$$Z_{I_1}=Z_I$$

$$Z_{I_2}=r_1^2r_2^2\ldots r_n^2Z_I$$

$$\theta=n\theta'$$

and the desired result has been achieved.

In the general case, the transformed network is not physically realizable because the transformation given by equation 1 leads to negative inductance or capacitance elements in some of the branches. There are, however, certain types of band-pass filter sections which may be transformed into sections with realizable elements for values of the transformation ratio not greater than a certain maximum. Furthermore in cascading sections it is sometimes possible to cancel negative elements in one section with positive elements in the mating section.

Two cases will be given as examples:

1. Case *a*. The transformation ratio varies from section to section. No negative circuit elements appear in the network.
2. Case *b*. The transformation ratio is the same for each section. No negative circuit elements appear in any interior branch of network.

Case a

Consider a network based on transformations of the symmetrical section of Figure 2. A special class of sections will be defined by the requirement that the reactances N_1 and N_2 may be physically subtracted from each other; that is a real constant $K>1$, independent of frequency, can be found such that

$$N_1-KN_2=0$$

Many standard sections, usually given in a different form, can be converted to the configuration of Figure 2 by the use of well-known circuit equivalences.

The first transformed section is also shown in Figure 2. From equation 1 its branches are given by:

$$Z_{A_1}=M_1+N_1-(r_1-1)N_2$$

$$Z_{B_1}=r_1^2(M_1+N_1)+r_1(r_1-1)N_2$$

$$Z_{C_1}=r_1N_2 \tag{2}$$

The maximum possible value of r_1 which will require no negative elements in the transformed section is, from the first of these equations

$$r_1=1+\frac{N_1}{N_2} \tag{3}$$

The branches of the second transformed section are

$$Z_{A_2}=r_1^2(M_1+N_1)-r_1^2(r_2-1)N_2$$

$$Z_{B_1}=r_1^2r_2^2(M_1+N_1)+r_1^2r_2(r_2-1)N_2$$

$$Z_{C_2}=r_1^2r_2N_2$$

When the two sections are cascaded, the second series branch of the network will have the impedance

$$Z_{B_1}+Z_{A_2}=2r_1^2(M_1+N_1)+2r_1^2N_2-r_1N_2-r_1^2r_2N_2$$

For a given r_1 the maximum value of r_2 which requires no negative element is that which reduces the sum of *N*-type impedances to zero, or

$$2r_1^2N_1+2r_1^2N_2-r_1N_2-r_1^2r_2N_2=0$$

$$r_2=2\left(\frac{N_1}{N_2}+1\right)-\frac{1}{r_1} \tag{4}$$

Similarly for the third section, the maximum transformation ratio r_3 is given by

$$r_3=2\left(\frac{N_1}{N_2}+1\right)-\frac{1}{r_2}$$

and in general for the *K*th section

$$r_k=2\left(\frac{N_1}{N_2}+1\right)-\frac{1}{r_{k-1}} \tag{5}$$

Let

$$x=\frac{N_1}{N_2}+1=r_1$$

Then

$$r_1=x$$

$$r_2=2x-\frac{1}{x}=\frac{2x^2-1}{x}$$

$$r_3=2x-\frac{x}{2x^2-1}=\frac{4x^3-3x}{2x^2-1}$$

Introduction of the Tschebyscheff polynomials[5] is convenient at this point. These are given by

$$T_1(x)=x$$

$$T_2(x)=2x^2-1$$

$$T_3(x)=4x^3-3x$$

The recurrence formula is

$$T_n(x)=2xT_{n-1}(x)-T_{n-2}(x)$$

From equation 5 it is seen that

$$r_1=T_1(x)$$

$$r_2=T_2(x)/T_1(x) \tag{6}$$

$$r_3=T_3(x)/T_2(x)$$

The extension to the general case

$$r_n=\frac{T_n(x)}{T_{n-1}(x)} \tag{7}$$

is obvious.

The over-all transformation ratio *r* of the network is the product of the transformation ratios of the individual sections

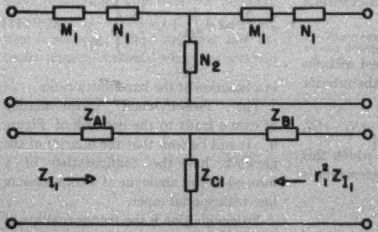


Figure 2 (left). Symmetrical section. N_1 of the series branch and N_2 of the shunt branch are understood to be the same type

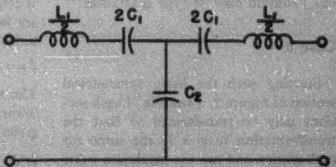
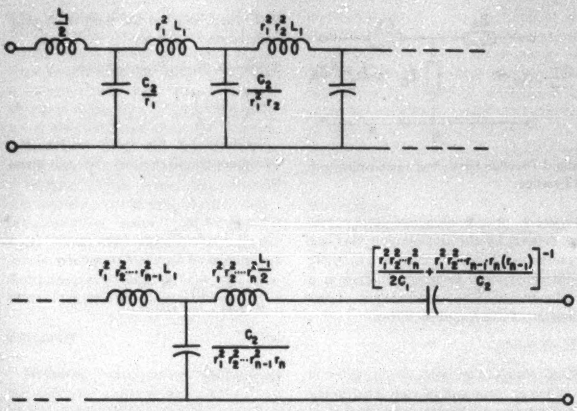


Figure 3 (right). Type III, section of Shea[4]



$\frac{L_1}{2}$
$r_1^2 L_1$
$r_1^2 r_2^2 L_1$
$\frac{C_2}{r_1}$
$\frac{C_2}{r_1^2 r_2}$
$r_1^2 r_2^2 \cdots r_{n-1}^2 L_1$
$r_1^2 r_2^2 \cdots r_n^2 \frac{L_1}{2}$
$\left[\frac{r_1^2 r_2^2 \cdots r_n^2}{2C_1} + \frac{r_1^2 r_2^2 \cdots r_{n-1}^2 r_n (r_n - 1)}{C_2}\right]^{-1}$
$\frac{C_2}{r_1^2 r_2^2 \cdots r_{n-1}^2 r_n}$

Figure 4. Result of "Tschebyscheff" transformation of case a

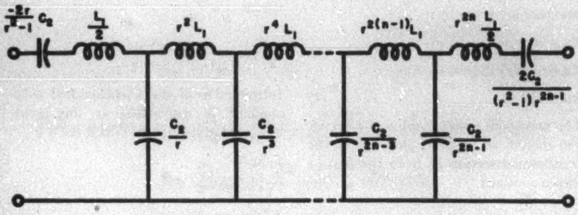


Figure 5. Result of transformation of case b type to section of Figure 3

$$r = r_1 r_2 r_3 \ldots r_r = T_1 \frac{T_2}{T_1} \frac{T_3}{T_2} \cdots \frac{T_n}{T_{n-1}} \qquad (8)$$

or

$$r = T_n(x) = T_n\left(1 + \frac{N_1}{N_2}\right)$$

The argument $x = (N_1/N_2) + 1$ is a constant depending on the parameters of the symmetrical section on which the network is based. It can usually be expressed in terms of the cutoff frequencies.

It will be noticed that the resulting network is economical, that is, the N-type elements have been made to disappear from the interior series branches so that all the series branches except the last (Z_{B_n}) contain only M-type reactances.

Case b

Starting with the basic symmetrical section of Figure 2, a cascade of such sections may be transformed so that the transformation ratio r is the same for each section and has the maximum value consistent with requiring no negative reactance elements in interior branches of the network. From equation 1 the branches of the first transformed section are given by

$$Z_{A_1} = M_1 + N_1 - (r-1)N_2$$

$$Z_{B_1} = r^2(M_1 + N_1) + r(r-1)N_2$$

$$Z_{C_1} = rN_2$$

The branches of the succeeding sections are, since r is constant,

$$Z_{A_2} = r^2 Z_{A_1} \qquad Z_{A_3} = r^4 Z_{A_1}$$

$$Z_{B_2} = r^2 Z_{B_1} \qquad Z_{B_3} = r^4 Z_{B_1}$$

$$Z_{C_2} = r^2 Z_{C_1} \qquad Z_{C_3} = r^4 Z_{C_1}$$

When the first two transformed sections are joined, the impedance of the interior series branch is

$$Z_{B_1} + Z_{A_2} = 2r^2(M_1 + N_1) + 2r^2 N_2 - rN_2 - r^3 N_2$$

The maximum value of r for which this branch contains no negative elements is given by

$$2r^2 N_1 + 2r^2 N_2 - rN_2 - r^3 N_2 = 0$$

$$r^2 - 2r\left(\frac{N_1}{N_2} + 1\right) + 1 = 0$$

Again letting

$$x = \frac{N_1}{N_2} + 1$$

we have

$$r = x + \sqrt{x^2 - 1} \qquad (9)$$

where the positive root has been chosen because we are looking for the maximum value of r. Similarly the conditions that no negative elements appear in the succeeding series branches of the transformed network lead in each case to equation 9.

It will be noted that the input series branch Z_{A_1} now contains a negative N-type reactance element, the interior series branches contain only positive M-type elements, and the output series branch contains a positive M-type and a positive N-type element. The negative element, being in an exterior branch, can under certain conditions be eliminated by adding a properly designed terminating section to the network.

Examples

Consider the type III$_1$ section of Shea[6] shown in Figure 3. This is a simple type of derived band-pass section with a peak of infinite attenuation at infinite frequency and with constant-K image impedance. The values of the reactance elements are

$$L_1/2 = R/(\omega_2 - \omega_1)$$
$$2C_1 = (\omega_2 - \omega_1)/\omega_1^2 R$$
$$C_2 = 2/(\omega_1 + \omega_2)R \qquad (10)$$

where ω_1 and ω_2 are the lower and upper cutoff (angular) frequencies and R is the mid-band image impedance. It will be seen that this section belongs to our special class with

$$M_1 = j\omega L_1/2$$

$$N_1 = 1/2j\omega C_1$$

$$N_2 = 1/j\omega C_2$$

The parameter

$$x = \frac{N_1}{N_2} + 1 = \frac{C_2}{2C_1} + 1 = \frac{\left(\frac{\omega_2}{\omega_1}\right)^2 + 1}{\left(\frac{\omega_2}{\omega_1}\right)^2 - 1} \qquad (11)$$

is a function of the bandwidth ratio.

The "Tschebyscheff" transformation of case a leads to the network of Figure 4. It will be seen that the interior of the network has the configuration of a lumped circuit analogue of a transmission line with special taper.

From equation 9 the transformation of

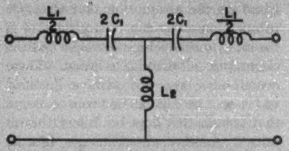


Figure 6. Compensating network for low impedance end of impedance transforming network

case *b*, again applied to the type III_1 section, leads to a transformation ratio

$$r=x+\sqrt{x^2-1}=\frac{\left(\frac{\omega_2}{\omega_1}\right)+1}{\left(\frac{\omega_2}{\omega_1}\right)-1} \tag{12}$$

The resulting network is shown in Figure 5. Its properties are easily studied since it is matched on an image basis. It has image parameters Z_1, $r^{2n}Z_1$ and $n\theta$ where Z_1 and θ are the image parameters of the basic 3-element symmetrical section.

One method of eliminating the negative capacitance is to add a section to the input (low-impedance) end of the network which contains in its series branch a series positive capacitance equal to or less than the magnitude of the negative capacitance. One such section, which has the same image impedance as the first section of the network, is the type III_2 section, shown in Figure 6. Its elements are

$$\frac{L_1}{2}=\omega_1 R/\omega_2(\omega_2-\omega_1)$$

$$2C_1=(\omega_2-\omega_1)/\omega_1\omega_2 R$$

$$L_2=(\omega_1+\omega_2)R/2\omega_1\omega_2$$

The capacitance $2C_1$ can be shown to be exactly equal to the magnitude of the negative capacitance of Figure 5.

The analysis can be carried out for Pi sections with similar results. Maximum transformations of transformable band-pass sections other than the 3-element sections also lead to economical structures. It is evident that the gain for a given bandwidth is greater per section with a case *b* structure than with a case *a*. Since an extra terminating section is always required to eliminate the negative capacitance at the end of a cascade of case *b* sections, case *a* sections are actually more economical when few sections are required while case *b* are more economical for a filter of many sections.

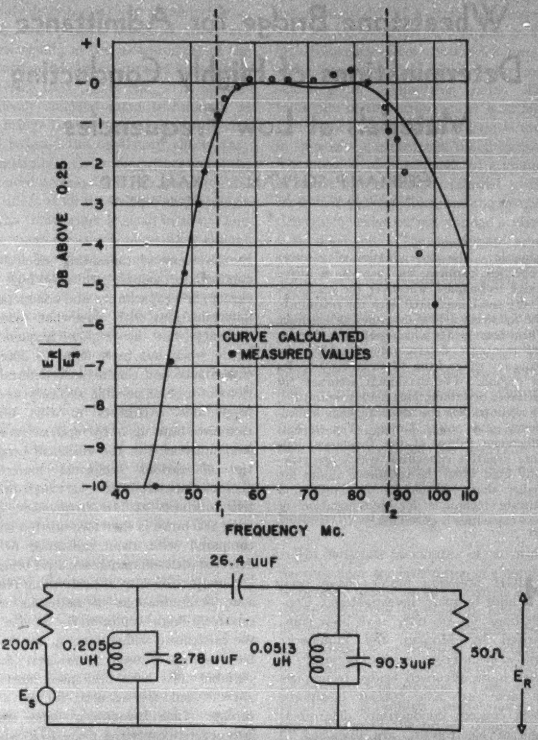


Figure 7. Bottom: Constants of a single-section type-a filter designed for 200-to 50-ohm transformations for the frequency range of television channels two through six. Top: Calculated and measured values of voltage transfer ratio for the filter below in the circuit shown

As an example of a simple single-section impedance transforming filter the case *a* Pi unit of Figure 7 was constructed and tested. Characteristics are shown in Figure 8. It may be seen that predicted properties are realizable to an excellent degree of approximation.

References

1. Transformer Band Filters, E. L. Norton. United States Patent *Number 1,681,544*, August 21, 1928.

2. Impedance Transformations in Band Pass Filters, A. S. Gladwin. *Wireless Engineer* (London, England), volume 20, November 1943, pages 540–47.

3. Extension of Norton's Method of Impedance Transformations to Band-Pass Filters, V. Belevitch. *Electrical Communication* (New York, N. Y.), volume 24, March 1947, pages 59–65.

4. Impedance Transformations in Four-Element Band Pass Filters, R. O. Rowlands. *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 37, November 1949, pages 1337–40.

5. See, for example, Die Mathematischen Hilfsmittel des Physikers (book), Erwin Madelung. Dover Publications, New York, N. Y., 1943, page 55.

6. Transmission Networks and Wave Filters (book), T. E. Shea. D. Van Nostrand Company, Inc., New York, N. Y., 1929, page 316.

No Discussion

Wheatstone Bridge for Admittance Determinations of Highly Conducting Materials at Low Frequencies

HERMAN P. SCHWAN
ASSOCIATE MEMBER AIEE

KARL SITTEL
NONMEMBER AIEE

Synopsis: The determination of reactive properties requires the use of sensitive bridge techniques, whenever the reactance is very small compared with the resistance. The following article describes an accurate Wheatstone bridge which permits resistance and reactance determinations within the frequency range from 10 to 200,000 cycles per second. The maximum accuracy for resistance determinations is one part in 10^{-7}; the accuracy for reactance depends on frequency and power factor. The various inductance values of the resistance coils cause negative capacities, which are especially high when the resistance values are smaller than 100 ohms. A simple and accurate technique for determination of these capacities is described.

HIGH precision bridges have been built by many investigators. Contributions up to 1926 have been summarized by Morgan and Lammert.[1] Jones *et al* published a series of papers in which more advanced bridge techniques are described and special questions, which concern conductivity determinations of electrolytes, are discussed.[2–8] Jones and Josephs achieved a maximum sensitivity of about one part in 10^{-6}. Their design proved to be very successful and further work in this field has been done with bridges of similar design. Another important contribution is Shedlovsky's investigation concerning proper shielding of the various bridge components.[9] More recently Luder[10] described an accurate conductivity bridge with very wide range, and Cole and Gross[11] constructed a wide-range capacitance-conductance bridge, which is useful up to 10 megacycles. Its main features are a special shifter network for balancing a hybrid-coil transformer arrangement, instead of an equal-ratio arm as introduced by Jones and Josephs and perfected by Campbell and Shakleton and Behr and Williams.[12] This 1 to 1 transformer automatically eliminates all undesirable unbalancing effects as caused by the generator or guard-electrodes; otherwise an additional Wagner ground arrangement[13] is necessary.

All these publications concern either the problem of measuring electrolytic conductance values with very high accuracy, or[11] capacitance and conductance determinations with somewhat reduced accuracy but at higher frequencies. Little work has been done to achieve capacitance and conductance determinations of highest possible accuracy at low frequencies. Attempts in this direction have been made by biphysicists who were interested in the electrical properties of various biological materials. Biological materials have very high dielectric constants at low frequencies, and their reactance is therefore not too small compared with their resistance to be detected with sufficiently sensitive bridges.

A major difficulty in performing reactance determinations of materials with relatively high conductivity is due to the existence of inductances in the variable resistors as used in bridges. Such variable units have been used most by authors and are essential for accurate tuning. Their inductance values simulate negative capacities $C=-L/R^2$ where $-C$ reaches values of more than 1,000 micromicrofarads in most resistance boxes when resistance R becomes smaller than 100 ohms. This high capacity depends strongly on the resistance box setting and directly influences the capacity determination. Knowledge of its value for each dial setting of the conductance box is therefore a necessity, and the accuracy with which these capacitances can be calibrated determines the accuracy of the capacitance measurements.

The determination of the reactive behavior of conductance and resistance boxes is a cumbersome task. Some biophysicists avoided this difficulty by use of a substitution method.[14,15] They replaced the biological material by an electrolytic solution whose total resistance was made equal to that of the biological material, either by a change of concentration or electrode distance. The resulting unbalance in capacity is eliminated by use of an additional variable capacitor parallel to the electrolytic solution. This substitution method is based on the assumption that electrode polarization either can be neglected or is the same in the biological sample and the electrolytic substitution solution. These assumptions are not always justified and special care has to be taken to assure that polarization does not affect the results.[16] Another disadvantage of the method is the difficulty of setting the resistance accurately. It is practically impossible to increase this accuracy beyond one part in 10^{-4} (0.01 millimeter for 10-centimeter electrode distance), which is insufficient for many investigations. Another arrangement is described by Schwan.[17] Here the determination of the electrical properties is achieved by use of a special direct reading bridge. The use of a simple slide-wire eliminates problems inherent in the use of a conductance box and effectively diminishes disturbing influences from the generator. However, the accuracy of this bridge is limited and reactance determinations under 1 kc are difficult even for biological materials with very high capacitance.

The bridge described in this article presents another solution to this problem. Its design is conventional and rather a simplification of the Jones and Joseph type bridge. However, amplification and tuned circuits have been used extensively to achieve as high a sensitivity as possible.

Description of Bridge

Figure 1 shows the wiring diagram of the bridge. The switches A and C permit throwing into each of the two grounded arms the unknown quantity X alone or in parallel to the resistance box 7 or the conductance box 6. The use of two resistance boxes would result in difficulties in adjusting to low resistance values accurately and for similar reasons two conductance boxes do not permit accurate tuning of high resistances. The use of a resistance box in one arm, together with that of a conductance box in the other arm has the advantage that it is possible to tune the resistive component of the bridge over the total range from 10 to 100,000 ohms with very high accuracy. This, together with the possibility of

---

Paper 53-127, recommended by the AIEE Electrical Techniques in Medicine and Biology Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted October 22, 1953; made available for printing December 29, 1952.

HERMAN P. SCHWAN is with the University of Pennsylvania, Philadelphia, Pa., and KARL SITTEL is with Franklin Institute, Philadelphia, Pa.

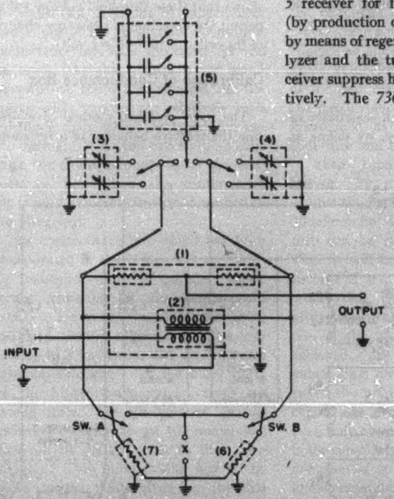
This article is a summary of Technical Report *NM00106201* of the Department of the Navy, Bureau of Aeronautics.

connecting the various capacitance boxes 3, 4, and 5 at will, makes the arrangement flexible. It can be used either for direct measurements or substitution determinations. For highest accuracy the following substitution method is used: At first, the unknown, X, is switched in parallel with the conductance box 6, the latter being set at its lowest conductivity. Balance is achieved by varying the resistance box setting with additional slight variations of the conductance box of necessary and by varying the capacitance boxes 3, 4, and 5. Then the unknown is disconnected and conductance 6 and capacitance 3 are increased by an amount equal to the unknown quantity to re-establish balance. The advantage of this method is that all static capacitances cancel out. One must take into account, however, the change of the capacitance of the conductance box, because of its different dial setting and the reactive properties of the connections to the unknown (inductance, stray field).

The bridge does not utilize a Wagner arm arrangement, since this would only complicate the procedure of measuring. The use of good electrostatically shielded transformers between oscillator and bridge prevents any disturbing influences from the oscillators. The substitution method already described is another safety factor, since each possible oscillator effect appears equally in the measurement of the unknown and the substituted impedance; therefore they cancel out. The electrostatically shielded input transformer of the bridge is part of the equal ratio arm and reduced the oscillatory effects very successfully. Grounding either one of the balanced terminals of the generator did not affect at all the balance of the bridge.

The transformer is small in size. Low resistances under 100 ohms draw enough current to reduce the output voltage of the transformer considerably. This was considered to be a minor disadvantage considering that sufficient magnetic shielding is more difficult to achieve with a larger transformer.

A Hewlett-Packard oscillator, type *205AG*, supplies the energy up to 20,000 cycles per second, and a General Radio oscillator, type *700A*, the higher frequencies. The output of the bridge is fed into a General Radio transformer, type *941A*, and stepped up 1 to 4 before it is applied to the input terminals of a Ballantine voltmeter, model *300*. This voltmeter serves as a preamplifier. Its output voltage finally is measured by a General Radio wave analyzer, model *736-A*, for frequencies under 16 kc, and by a Radio Corporation of America *RAK-5* receiver for frequencies above 15 kc (by production of an audible beat signal by means of regeneration). The wave analyzer and the tuned radio-frequency receiver suppress harmonics and noise effectively. The *736-A* wave analyzer only has a bandwidth of 4 cycles per second. The *RAK-5* receiver, in addition to its high frequency circuits, utilizes a special audio-frequency filter which helps considerably in suppressing noise. The total gain of this arrangement is so high that noise occurs when full gain is used in spite of the presence of tuned circuits. The noise level corresponds to a voltage of about 0.03 microvolt, referred to the input of the preamplifier, and the smallest detectable bridge output voltage is about 0.01 microvolt. The maximum sensitivity of the bridge, referred to an input voltage of 1 volt, is therefore given by a ratio of better than 1 to 10^{-7}. All the various resistances, capacitors, and transformers are shielded, and the total bridge is itself built into a metal enclosure. The generators and receivers are also shielded and mounted on a relay rack. This, together with proper grounding by low inductive copper bands, reduces the direct influence from the generator to the receiver to an unnoticeable amount.



(1), (2)	SHIELDED RATIO BOX, LEEDS & NORTHRUP CAT. NO. 1553 ELECTROSTATICALLY SHIELDED INPUT TRANSFORMER
(3), (4)	PRECISION CONDENSER, GENERAL RADIO TYPE 722 DQ
(5)	DECADE CAPACITOR, GENERAL RADIO TYPE 219 K
(6)	SHIELDED RESISTANCE BOX, LEEDS & NORTHRUP CAT. NO. 4749
(7)	SHIELDED CONDUCTANCE BOX, LEEDS & NORTHRUP CAT. 4764-S

Figure 1. Bridge circuit

Sensitivity of Bridge

The absolute accuracy of a bridge depends upon the interaction of a number of factors. More or less constant errors result from the mechanical limits inherent in setting the dials; variable errors result from the variation in the accuracy of calibration of the electrical components at different dial settings. The balance sensitivity is limited, furthermore, by the function of the bridge components, by the amplifiers applied to the bridge output, by the applied generator voltage, and by the inherent noise level. Frequency precision is another factor requiring meticulous control in many instances. It is achieved at lower frequencies by comparison of the power-line frequency with an oscillograph and at higher frequencies by use of harmonics, as produced by a square-wave generator whose fundamental has been compared with the tenth harmonic of the power frequency.

The upper limit of bridge sensitivity depends on the mechanical limitation of dial settings, and is equal to the ratio of the smallest possible dial change to the total quantity to be measured. The resistance box has a total maximum resistance of 11,11.11 ohms with the smallest possible change being equal to 0.01 ohm. The conductance box incorporates a slide-wire which permits conductance changes of less than 0.01 micromho and has a maximum conductance of 111,111.11 micromhos. The common use of both boxes in different bridge arms permits the tuning of resist-

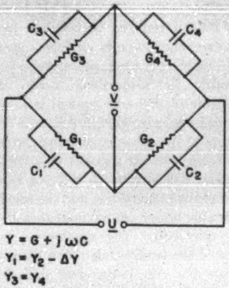


Figure 2. Definition of terminology

ances to exactly 10^{-6} ohm for high resistance values and better than 10^{-7} ohm for low values. The variable capacitors are calibrated within 0.1 micromicrofarad in their 25 to 110-micromicrofarad range and exactly within 0.1 micromicrofarad up to 1,100 micromicrofarads. They can be set with the same precision.

The absolute electrical accuracy of the resistance and conductance box is 0.1 per cent. It should be pointed out that the use of a bridge, which permits, let us say, a sensitivity of one part in a million does not mean that determinations with this accuracy can be carried out. It permits, however, detection of changes of this order of magnitude.

If an equal ratio arm bridge has admittances Y_3 and Y_4 equal (see Figure 2) and if further the output voltage V is measured with a voltmeter of high impedance as compared to the bridge impedances, then the following relationship holds

$$\frac{V}{U}=\frac{1}{4}\,\frac{\Delta Y/Y}{1-\frac{1}{2}(\Delta Y/Y)} \tag{1}$$

or

$$\frac{V}{U}=\frac{1}{4}\,\frac{\Delta Y}{Y} \tag{2}$$

when only small deviations ΔY from the correct bridge balance $\Delta Y=0$ are considered. Measuring the output with an amplifier, which registers only amplitude, but not phase, yields

$$V=\frac{U}{4}\sqrt{\frac{(\Delta G)^2+(\omega\Delta C)^2}{G^2+(\omega C)^2}} \tag{3}$$

where

U and V are amplitudes of voltages V and U
ΔG and ΔC are conductance and capacitance deviations from point of balance as defined by $\Delta Y=\Delta G+j\omega C$, $Y=G+j\omega C$
ω is angular frequency

Two special cases require further consideration: One is where the capacitances C_1 and C_2 are equal, for example, when the capacitances are correctly balanced but the conductances are slightly out of balance. The second case assumes correct balance of the conductance values but slight unbalance of the capacitances. In the first case where $\Delta C=0$, the condition for perfect capacitive balance is

$$V_G=\frac{U}{4}\,\frac{\Delta G}{\sqrt{G^2+\omega^2C^2}} \tag{4}$$

In the second case, where $G=0$, for example, the case of perfect conductive balance

$$V_C=\frac{U}{4}\,\frac{\omega\Delta C}{\sqrt{G^2+\omega^2C^2}} \tag{5}$$

The comparison of both results shows that the smallest noticeable conductance change ΔG_0 is equal to the smallest noticeable admittance change $\omega\Delta C_0$. This can be expressed as follows

$$\frac{\Delta G_0/G}{\Delta C_0/C}=\frac{\omega C}{G} \tag{6}$$

which means that the ratio of the smallest noticeable relative changes of capacitance and conductance equals the loss tangent of the material to be investigated. This value is very high at low frequencies for all substances which are good conductors. Consequently, the smallest noticeable conductance change is much smaller in all such cases than the smallest noticeable relative capacity change. This necessitates a very high sensitivity for the conductance, in order to be able to determine the susceptance with some reasonable degree of accuracy.

To characterize the high sensitivity of the bridge by high values, we define as conductive and capacitive sensitivity the following quantities for a bridge input voltage $U=1$ volt

$$\epsilon_G=\frac{G}{\Delta G_0};\quad \epsilon_C=\frac{C}{\Delta C_0} \tag{7}$$

where ΔG_0 and ΔC_0 are the smallest measurable conductance and capacitance changes. Introducing this into equations 4 and 5, it is possible to express the sensitivities in terms of the smallest detectable voltage V_6 as follows

$$\frac{1}{\epsilon_G}=4V_0\sqrt{1+\left(\frac{\omega C}{G}\right)^2};\qquad \frac{1}{\epsilon_C}=4V_0\sqrt{1+\left(\frac{G}{\omega C}\right)^2} \tag{8}$$

The conductive sensitivity obviously is independent of the frequency, as long as $(\omega C/G)<0.1$ and equal to $1/(4V_0)$. Under the same assumption, the capacitive sensitivity increases in proportion with ω to a maximum value of $1/(4V_0)$ at higher frequencies.

Determinations of the smallest voltage detectable by the amplifier section have been carried out over the frequency range of 20 to 200,000 cycles per second. It was found that this voltage is relatively independent of the frequency between 100 and 200,000 cycles per second and has a value of about 0.02 microvolt. Introducing these figures in equation 8 shows that these threshold voltage values permit the detection of resistance changes of the order of 10^{-7} ($U=1$ volt).

Calibration of Conductance Box

The most practical method of expressing the reactive behavior of a resistance

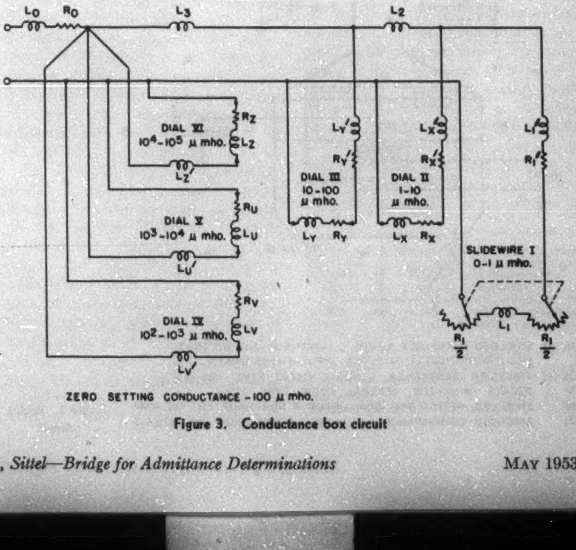
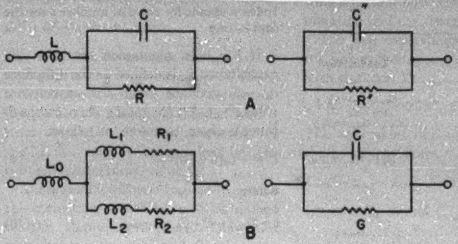


Figure 3. Conductance box circuit

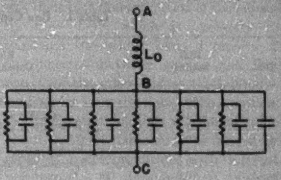
Figure 4 (left). Transformation of two conductance box circuits



(A) $R \approx R''$
$C \approx C'' - (L/R^2)$

(B) $G[1+(\omega C/G)^2] \approx G_1+G_2$
$C \approx C_1 + C_2 - 2L_0G_1G_2$

Figure 5 (right). Equivalent circuit of conductance box



box, which has all its variable decade resistors in series, is to transform the reactive properties of each resistor into an inductance in series with the resistor. Reactive properties of a conductance box, however, with all its decade units in parallel are conveniently expressed by capacitances in parallel with the conductance. If this is done, the total reactance is, in both cases, given by the sum of the inductances or capacitances of all the dials in series or in parallel respectively.

For capacitive measurements, where the reactive component is tuned by capacitors in parallel with the variable resistor units, it is more convenient to have the reactive properties of the variable box expressed in terms of capacitance. This is one of the reasons for choosing a conductance box as a substitute for the unknown. Another reason is the small influence of contact resistances on the tuning while varying the relatively high resistance dials which are used for the fine tuning. It is only necessary to calibrate the reactive properties of the conductance box when the conductance box is used as a substitute as described in the foregoing.

The conductance box which has been incorporated in the bridge has six variable resistance units in parallel, which permit variation of the conductance from 0 to 1.0 to 10.0 to 1,000.0 to 10,000, and 0 to 100,000 micromhos. The first unit up to 1 micromho uses a continuous variable slide-wire. All the other units for the higher conductance values are decade units. The conductance box has an initial conductance of 100 micromhos when all its dials are set to their zero position. This initial value has to be added to the value given by the dial setting. The box permits, therefore, an absolute variation from 100 to 111,211 micromhos. Its usable range, however, is much greater with the aforementioned substitution method, where the initial 100-micromho value is also present across the unknown and cancels out.

The principle of the calibration of the reactive properties of the conductance box can be described as follows: Suppose the capacitance of the slide-wire dial for its highest setting, 1 micromho is known to be x with reference to its zero setting. The bridge may be balanced for this 1-microhmo setting. By setting this dial back to zero and setting the next dial (range 0 to 10 micromhos) to its first position of 1 micromho no disturbance occurs in the resistive part of the bridge balance. However, the capacitive balance has to be re-established, because the 1-micromho setting of the 1 to 10-micromho dial has a different capacity than the slide-wire. The change in capacitance, which is required for re-establishing balance as added to x, gives the capacitance value for the 1-micromho setting of the new dial. The total capacity of the conductance box, for a setting of 2 micromhos, as achieved by setting the slide-wire and the 1 to 10-micromho dial to their 1-micromho reading is then known. By exchange of this setting against the 2-micromho setting of the 1 to 10-micromho dial the 2-micromho setting can be calibrated in terms of x. Thus it will be possible to calibrate by successive steps the total 1 to 10-micromho dial. By application of these same procedures to the other dials, the calibration of the whole box in terms of x is achieved. All the various capacitance values thus obtained are differences between the total box capacity and the box capacity with all dials set to zero. However, only these differences are of interest in the substitution procedure. The determination of x requires one standard of known resistance and reactance which can be exchanged with the conductance box. This standard should have a relatively high conductance. If its value is low, the continuous repetition of the step-up procedure will apply an initially small error in the determination of x always in the same direction. This would lead to considerable errors in the capacitance values for the high conductance settings of the box.

Various distributed inductances effect the step-up procedure, in spite of their low values, and require a closer analysis. Figure 3 shows the total circuit of the conductance box. The slide-wire dial and the other five decade dials, x, y, z, u, and v are connected with wires and copper bands of low resistance and inductance. All resistors are assumed to be R-L series circuits. They are mutually connected to the bridge by the connection $R_0 - L_0$. Distributed capacitances between the various connections are omitted since they cancel out in the substitution procedure. To better understand the effect of the various elements, the following basic circuits as illustrated in Figure 4 should be discussed. Figure 4(A) shows an inductance L in series with an $R-C$ parallel combination and its equivalent parallel combination with the same impedance. The inductance, L, in the case of the conductance box, is caused by the lead L_0 which connects the box with the bridge. The products ωCR are in this case so small that they can be neglected as compared to 1. Then the effect of L can be described by a frequency-independent negative capacity $-L/R^2$. For the same reason, the deviation between R and R'' can be neglected. Figure 4(B) is somewhat more complicated. It shows two inductive resistors in parallel and in series with an inductance L_0. The equation which describes how the total capacitance C of this system can be expressed is obtained under the assumption that the susceptance ωC is small compared to the resistive conductances, G.

The application of the aforementioned circuits permits analysis of the total circuit of the conductance box. The inductance values L_2 and L_3 appear only in line with the variable resistors R_1, R_x, and R_y and their inductance values L_1, L_x, and L_y. All these resistors are higher than 600 ohms and the inductance values L_2 and L_3, represented by relatively short connections smaller than 0.1 microhenry.

Table I. Box Capacitances for Dial u

Dial	Dial Setting	Box Capacitance	Conductance
$u-1$	10	$C_{(u-1),10}{}^B = a$	$G_{(u-1),10} \approx G_{u,1}$
u	1	$C_{u,1}{}^B = a + \Delta C_{u,1}{}^B$	$G_{u,1}$
u $u-1$	1 10	$C_{u,1}{}^{B'} = 2a + \Delta C_{u,1}{}^B - 2L_0G_{u,1}G_{u,1}$	$G_{u,1} + G_{u,1} \approx G_{u,2}$
u	2	$C_{u,2}{}^B = 2a + \Delta C_{u,1}{}^B + \Delta C_{u,2}{}^B - 2L_0G_{u,1}G_{u,1}$	$G_{u,2}$
u $u-1$	2 10	$C_{,u2}{}^{B'} = 3a + \Delta C_{u,1}{}^B + \Delta C_{u,2}{}^B - 2L_0G_{u,1}(G_{u,1} + G_{u,2})$	$G_{u,2} + G_{u,1} \approx G_{,u3}$
u	3	$C_{u,3}{}^B = 3a + \Delta C_{u,1}{}^B + \Delta C_{u,2}{}^B + \Delta C_{u,3}{}^B$	$G_{u,3}$
.			.
.			.
.			.
u	n	$C_{u,n}{}^B = na + \sum\limits_{\nu=1}^{\nu=n} \Delta C_{u,\nu} - 2L_0G_{u,1} \sum\limits_{\nu=1}^{\nu=n-1} G_{u,\nu}$	$G_{u,n}$
.			.
.			.
.			.
u	10	$C_{u,10}{}^B = 10a + \sum\limits_{\nu=1}^{\nu=10} \Delta C_{u,\nu} - 2L_0G_{u,1} \sum\limits_{\nu=1}^{\nu=9} G_{u,\nu}$	$G_{u,10}$

$a = x, y, z, u, v$ according to $u = 2, 3, 4, 5, 6$

Index u stands for the dial number ν for the dial setting

The capacitive effect of these inductances is smaller than 0.3 micromicrofarad and in the majority of cases even smaller than 0.1 micromicrofarad. It will be neglected, therefore. On the other hand, the inductance L_0 may be in series with resistors as small as 10 ohms and has been found to be 0.17 microhenry. It can cause a negative capacity of the order of 1,000 micromicrofarads and definitely must be considered.

All connections within the box are relatively short and their resistances are so small, compared with the resistance of the variable units, that they can be neglected. The same holds true for the resistance R_0 of the leads which connect the box with the bridge. However, R', R_x', and R_y' are resistors in series with the lead wires. The inductances L_1, L_x, and L_y can be lumped together with the inductance values of the variable units and may always be expressed as capacities in accordance with the equations as given in Figure 4. Thus, it has been shown that the total conductance box circuit can be replaced by the circuit which is shown in Figure 5.

The terminology used to express all quantities is as follows: More than one dial set different from zero is called total gross-box capacitance. Only one dial set different from its zero setting is called gross-box capacitance. These two quantities correspond to certain capacitances between points B and C, which are called total-gross-dial-capacitance and gross-dial-capacitance. All dials set at zero result in the residual box capacitance across AC and the corresponding residual dial capacitance across BC. When these residual values are subtracted from the first defined quantities, the differences are called box capacitances or total box capacitances respectively. The symbols which will be used for the different quantities are as follows:

Capacitance as appearing across dial resistors *BC*:	**Symbol**
Gross dial capacitance	$c_{u,\nu}$
Total gross dial capacitance	c
Dial capacitance	$C_{u,\nu}$
Total dial capacitance	C
Residual dial capacitance	C_0
Capacitance as measured across box terminals *AC*:	
Gross box capacitance	$c_{u,\nu}{}^B$
Total gross box capacitance	c^B
Box capacitance	$C_{u,\nu}{}^B$
Total box capacitance	C^B
Residual box capacitance	$C_0{}^B$
Conductance	$G_{u,\nu}$
Total conductance resulting from a specified setting of all six dials less the residual conductance	G
Residual conductance (sum of the residual) values of the different dials in zero position	G_0

It has been mentioned that the dial capacitance C_u is defined as the difference $c_{u,\nu} - C_0$ and the total box capacitance C^B as $c^B - C_0{}^B$. The total gross box capacitance is given, therefore, as follows:

$$c^B = -L_0(G+G_0)^2 + c \tag{9}$$

where

$$c = \Sigma c_{u,\nu} = C + C_0 \tag{10}$$

This equation is a direct consequence of the equations depicted in Figure 4(B), when it is assumed that the products ωRC can be neglected. The residual box capacitance is obtained when all dial settings are zero

$$C_0{}^B = -L_0G_0{}^2 + C_0 \tag{11}$$

The difference between the total gross box capacitance, for any arbitrary dial setting, and the residual value $C_0{}^B$ gives the total box capacitance C^B. Applying the previously described substitution method

$$c^B - C_0{}^B = C^B = C - L_0G^2 - 2L_0GG_0 \tag{12}$$

By this equation the capacitance changes are determined when the conductivity is increased by an amount ΔG

$$C^{B'} - C^B = \Delta C - L_0\Delta G^2 - 2L_0\Delta G(G+G_0) \tag{13}$$

ΔC is the change in total dial capacitance due to the conductance change ΔG. The conductance ΔG alone, as applied to the box when its initial setting is zero, would cause a total box capacitance

$$\Delta C^B = \Delta C - L_0\Delta G^2 - 2L_0\Delta GG_0 \tag{14}$$

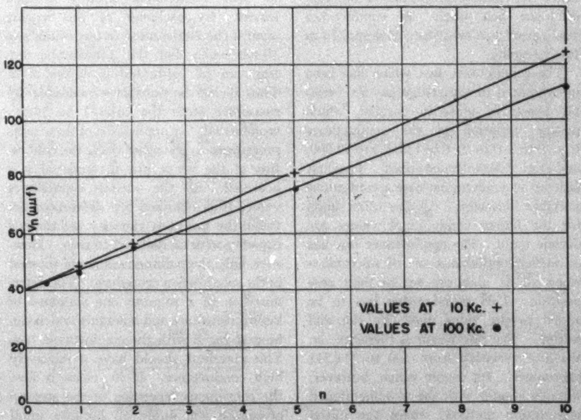


Figure 6. Capacitance V as function of conductance

Equation 13 can be rewritten

$$C^{B'} - C^{B} = \Delta C^{B} - 2L_0 \Delta G G \qquad (15)$$

The total box capacitance change due to a change of the conductance G is, therefore, identical with the total box capacitance due to a change of the dial setting from zero to ΔG when corrected for $-2L_0 G \Delta G$, and is independent of the residual conductance value G_0.

These results will now be used to discuss the corrections which have to be applied to the step-up procedure. Only box capacitances, as measured between the terminals of the conductance box, can be observed while performing the calibration of the box. The box capacitances for the $\nu = 10$ settings are used as build-up units in the step-up procedure. For abbreviation these values shall be called x, y, z, u, v with u being 2, 3, 4, 5, 6 respectively. Table I shows for one dial how the box capacitances can be related to the initial build-up value $a = C^{B}_{n-1,10}$. Its values are derived by successive use of the step-up principle and corrections in accordance with equation 15. The $\Delta C_{u,\nu}{}^{B}$ values are the observed changes of the box capacitance which are required for re-establishing bridge balance while exchanging the setting $G_{u-1,10} = G_{u,\nu}$ against the settings $G_{u,\nu+1}$.

If the effect of L_0 did not exist, the following relationship for the capacitance associated with the conductance setting $G_{u,\nu}$ would be valid for $\nu = n$

$$C_{u,n}{}^{B} = na + \sum_{\nu=1}^{n} \Delta C_{u,\nu}{}^{B} \qquad (16)$$

Including effects caused by L_0, we obtain, however

$$C_{u,n}{}^{B} = na + \sum_{\nu=1}^{n} \Delta C_{u,\nu}{}^{B} - 2L_0 G_{u,1} \sum_{\nu=1}^{n} G_{u,\nu} \qquad (17)$$

The simple-step-up procedure as expressed by equation 16, therefore, has to be corrected for $-2L_0 G_{u,1}$. Equation 16 simplified, includes L_0 values different from zero when we replace the build-up value by a value a_n which is defined by

$$n(a_n - a) = -2L_0 G_{u,1} \sum_{\nu=1}^{n-1} G_{u,\nu} \qquad (18)$$

and which is dependent on the dial setting as indicated by the indices n. If $G_{u,n} = nG_{u,1}$ we can replace $\sum_{\nu=1}^{n-1} G_{u,\nu}$ by $\frac{n}{2}(n-1)G_{u,1}$ and obtain

$$a_n = a - (n-1)L_0 G_{u,1}{}^{2} \qquad (19)$$

This equation has been used for a determination of L_0 as described later on.

The dial capacitance A equivalent to the box capacitance a can be determined with equation 12

$$a = A - L_0 G_{u,1}{}^{2} - 2L_0 G_{u,1} G_0 \qquad (20)$$

The equivalent dial capacitance $C_{u,n}$ to the box capacitance $C_{u,n}{}^{B}$ is

$$C_{u,n}{}^{B} = C_{u,n} - L_0 G_{u,n}{}^{2} - 2L_0 G_{u,n} G_0 \qquad (21)$$

Introducing A and $C_{u,n}$ in equation 17

$$C_{u,n} = nA + \sum_{\nu=1}^{n} \Delta C_{u,\nu}{}^{B} \qquad (22)$$

This result shows that the simple step-up equation 16 remains correct when applied to the dial capacitance values regardless how big L_0 and G_0 are. The validity of equation 22 is evident when we return to a brief discussion of the circuit in Figure 5 (A). It is obvious that the simple step-up method is correct when applied to the dial capacitances as measured between points B and C. The differences ΔC which are required for rebalancing the bridge while exchanging $G_{u-1,10} + G_{u,\nu}$ against $G_{u,\nu+1}$ are here, of course, referred to the points B and C. However, the difference between box capacitance and dial capacitance is only dependent on L_0 and $G + G_0$ and independent of the dial combination by which the total conductivity is achieved. Therefore this difference cancels out when we determine the differences in C while exchanging the position $u+1,10+u,\nu$ against $u,\nu+1$. This makes it possible to replace the sum $\Sigma \Delta C_{u,\nu}$ by $\Sigma \Delta C_{u,\nu}{}^{B}$ and yields equation 22. The final result of the calculations, as presented in equation 22, is evident. The calculations, however, are still of value, since they provide methods for the determination of the fundamental constants A and L_0.

Several standards have been used for the calibration of the reactive behavior of the conductance box. Four standard resistors, type *663*, were purchased from General Radio Company. These are small wire-wound resistors with a special construction which minimizes magnetic fields and reduces their inductance to a very small value, determined by the manufacturer. Two additional standard resistors were built in our laboratory and consist of thin manganin wire, which is stretched in the center of a brass cylinder. The inductance of this arrangement has been determined by calculation. The distributed inductance values of various leads which connect the variable units and terminals with the bridge points A and C have been determined with the greatest possible accuracy. This was done by a determination of the capacity $C = -L/R^2$ which is necessary to rebalance the bridge when a resistor R is connected between ground and first one, then the other end of the inductive lead under investigation. The accuracy in the L determinations was found to be about ±0.0001 microhenry as proved by repeated L determinations with different resistors R. All standards had different resistance values. The General Radio resistors had 10, 20, 50 and 100 ohms and our standards 25.5 and 250 ohms.

The following procedure has been chosen for calibration. The standards were at first in parallel with the conductance box, the latter with all dials at zero. When the standard was removed, the conductance of the box increased to the value of the standard and the capacitance determined which was necessary to re-establish balance. From this capacitance the expression L/R^2 was subtracted, where L is the combined inductance of the standard and the leads which connect it to the conductance box terminal of the bridge. The difference is then due to the conductance box and its L_0 up to the point where the conductance box and the standard are linked together. The result has been expressed in terms of a_n by utilizing the equations for the simple step-up procedure and all values which had been determined before. Figure 6 shows the results of these determinations. It is obvious that $a_n = \nu_n$ changes linearly with the conductance which was present during the calibration. This is in agreement with equation 19 and makes it possible to determine L_0. Its value, as given by the slope of the curve, is equal to 0.084 microhenry at 10 kc and 0.071 microhenry at 100 kc. Subsequent use of equations 19 and 20 for the case $n = 10$ made it possible to determine all quantities $a{:}(x, y, z, u, \text{and } v)$ and corresponding dial values $A{:}(X, Y, Z, U, V)$. After that equation 22 was used for determination of all other dial values $C_{u,\nu}$. There is no value in calculating all corresponding box capacitances, since the total box capacitance for any arbitrary dial setting is not equal to the sum of all the single box capacitances of the various dials. However, the sum of all the single dial capacitances is equal to the total dial capacitance. It is therefore more practical to add up all dial values and to apply the correction $-L_0 G^2 - 2L_0 G G_0$, equation 12, afterwards for each measurement performed with the bridge. It will be found that the correction, especially its first term, can be very often neglected in cases where either relatively high capacitances are observed or when the conduc-

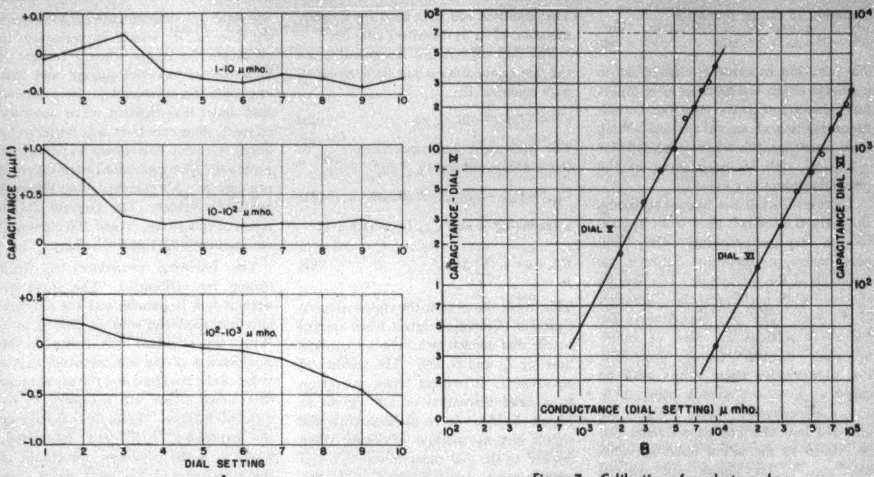


Figure 7. Calibration of conductance box

tivity of the sample is low.

Figure 7 shows the results of the calibration. Calibrations have been performed at 3, 10, 30, and 100 kc. The results agree within the accuracy of the measurements for all dials except the last, one. In this case, the results as given in Table II are obtained. The fact that the capacitance changes with the square of the dial setting for the dials V and VI is explained when we assume that the low resistance coils of these dials do not add materially to the inductance of the various leads which connect these resistors separately with the box terminals. It is apparent further that this is no longer the case for the dials II, III, and IV.

If the inductance were completely concentrated in the resistor and proportional to *R*, equation $C=L/R^2$ would result in a capacitance proportional to the first power of the conductance. Curve IV represents a case where a nearly constant value, *C*, as given by dial III, changes at first to a linear behavior and then approaches a square behavior as a function of the conductance. This total *C* has superimposed on it fluctuations due to small irregular changes of both static capacity and conductance.

Table II. Dial Capacitances in Micromicrofarads of Dial VI of Conductance Box at Various Frequencies

n	3 kc	10 kc	30 kc	100 kc
1	−36	−35	−36	−36
2	−137	−134.5	−135	−134
3	−285	−279	−278	−275
4	−500	−502	−497	−492
5	−709	−699	−691	−681
6	−960	−948	−935	−921
7	−1,470	−1,451	−1,435	−1,415
8	−1,860	−1,837	−1,816	−1,790
9	−2,220	−2,184	−2,156	−2,120
10	−2,850	−2,800	−2,763	−2,710

It was suspected that the change of the dial capacitances of dial VI was due to skin effect in the leads to the resistive elements. If this is true, all capacitance changes for the various dial settings, as observed between two of the various frequencies, must be explained by the same inductance change. Table III gives the change of *L* due to skin effect and as calculated by the equation $\Delta C=-\Delta L/R^2$. The ΔC values are the differences between the dial capacitances as determined at 3, 10, and 30 kc and the dial values at 100 kc. The ΔL values are indeed constant for each frequency within the accuracy of determination. Further proof was obtained by analytical study of the inductance of a copper bar of equal dimensions as those used as connecting leads inside the conductance box. This led to a predicted change of 0.005 microhenry between 10 and 100 kc which is considered in good agreement with the experimental result of 0.007 microhenry. The calculation further predicted that the frequency range of the *L* change due to skin effect should be about 0.003 microhenry between 3 kc and 100 kc, approximating the ultimate values at very low and high frequency within 1 per cent. Thus it was possible to say that the calibration of the box has been performed for the total frequency range of the bridge.

The conductance box calibration becomes more and more frequency-dependent when values ωL can no longer be neglected compared to the various resistors, *R*. Introducing *R* equal to 10 ohms, its *L* of about 0.3 microhenry yields a frequency of about 600 kc for a 1-per-cent deviation in *C*.

The only dial which could not be calibrated with the step-up method was the slide-wire dial I. However, its final capacitance for its highest setting of 1 micromho is known from the step-up calibration to be equal to +0.20 micromicrofarad. It was found experimentally that the capacitance change which is necessary to rebalance the bridge, when the dial setting is decreased from 1 to 0 micromhos, is identical with this value within 0.02 micromicrofarad. The change

Table III. L_0 as Function of Dial Setting and Frequency

	3 kc		10 kc		30 kc	
n	ΔC, Micro-micro-farads	ΔL, Micro-henry	ΔC, Micro-micro-farads	ΔL, Micro-henry	ΔC, Micro-micro-farads	ΔL, Micro-henry
2	3		2		2	
3	10	0.011	4		3	
4	17	0.011	10	0.006	5	0.003
5	28	0.011	18	0.008	10	0.004
6	39	0.011	27	0.007	14	0.004
7	55	0.011	36	0.007	20	0.004
8	70	0.011	47	0.007	26	0.004
9	100	0.012	64	0.008	36	0.004
10	140	0.014	90	0.009	53	0.005

in the resistance box setting, which is required for re-establishing bridge balance, does not affect the capacitance setting. (A study of the resistance box showed that its contribution to the total bridge reactance was small compared to that of the conductance box as long as its resistance remained high. It becomes stronger, however, than the conductance box effects for lower resistance values.) This fact made it possible to determine capacitances of the slide-wire dial over its range by slight variations of the resistance box and the slide-wire dial I at the same time. The result of this calibration shows that the slide-wire capacity increases continuously from 0 to 0.18 micromicrofarad.

References

1. Factors Influencing the Accuracy of Measurements of the Electrical Conductance of Liquids and Solutions—II, **J. L. R. Morgan, O. M. Lamnert.** *Journal*, American Chemical Society (New York, N. Y.), volume 48, 1926, page 1220.

2. The Measurement of the Conductance of Electrolytes—I, **G. Jones, Josephs.** *Journal*, American Chemical Society (New York, N. Y.), volume 50, 1928, page 1049.

3. The Measurement of the Conductance of Electrolytes—II, **G. Jones, G. M. Bollinger.** *Journal*, American Chemical Society (New York, N. Y.), volume 51, 1929, page 2407.

4. The Measurement of the Conductance of Electrolytes—III, **G. Jones, G. M. Bollinger.** *Journal*, American Chemical Society (New York, N. Y.), volume 53, 1931, page 411.

5. The Measurement of the Conductance of Electrolytes—IV. *Ibid.*, page 1207.

6. The Measurement of the Conductance of Electrolytes—V, **G. Jones, B. C. Bradshaw.** *Journal*, American Chemical Society (New York, N. Y.), volume 55, 1933, page 1780.

7. The Measurement of the Conductance of Electrolytes—VI, **G. Jones, Sch. M. Christian.** *Journal*, American Chemical Society (New York, N. Y.), volume 57, 1935, page 272.

8. The Measurement of the Conductance of Electrolytes—VII, **G. Jones, D. M. Bollinger.** *Ibid.*, page 280.

9. A Screened Bridge for the Measurement of Electrolytic Conductance—I, II, **T. H. Shedlovsky.** *Journal*, American Chemical Society (New York, N. Y.), volume 52, 1930, page 1793.

10. A Variable Ratio-Arm Conductivity Bridge, **W. Luder.** *The Review of Scientific Instruments* (New York, N. Y.), volume 14, 1943, page 1.

11. A Wide Range Capacitance-Conductance Bridge, **R. H. Cole, P. M. Gross.** *The Review of Scientific Instruments* (New York, N. Y.), volume 20, 1949, page 252.

12. The Campbell Shackelton Shielded Ratio Box, **J. Behr, A. J. Williams.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), June 1932.

13. Zur Messung Dielektrischer Verluste mit der Wechselstrombruecke, **K. W. Wagner.** *Elektrotechnische Zeitschrift* (Berlin, Germany), volume 32, 1911, page 1001.

14. The Electrical Impedance of Haemolized Suspensions of Mammalian Erythrocytes, **H. Fricke, H. J. Curtis.** *Journal of General Physiology*, volume 18, 1935, page 821.

15. Electrical Impedance of Hipponoe Eggs, **K. S. Cole.** *Journal of General Physiology*, volume 18, 1935, page 821.

16. Die temperaturabhaengigkeit der Dielektrizitaetskonstante von Blut bei Niederfrequenz, **H. Schwan.** *Zeitschrift für Naturforschung* (Wiesbaden, Germany), volume 3, part B, 1948, page 361.

17. Elektrodenpolarisation und ihr Einfluss auf die Bestimmung dielektrischer Eigenschaften von Fluessigkeiten und biologischem Material. *Ibid.*, volume 6, part B, 1951, page 121.

No Discussion

High-Power Industrial Vacuum Tubes Having Thoriated-Tungsten Filaments

R. B. AYER
NONMEMBER AIEE

Synopsis: Contrary to a long-standing belief, thoriated-tungsten filaments can be used as electron emitters in high-power vacuum tubes with a reliability equal to that established by low-power types. High-power types may be defined as those having unmodulated class *C* plate voltage ratings of greater than 5,000 volts. The use of thoriated-tungsten filaments in high-power types has been made possible by the use of new and improved materials and processing techniques. In the radio broadcast field, performance of these high-power tubes has been excellent. In the industrial field, tubes having thoriated-tungsten filaments are continuing to set new records after as much as 22 months of uninterrupted full-load service at plate voltages as high as 17,000 volts. Such performance, in addition to the lower cost of equipment components and lower cost of operation due to reduced filament-power requirements, makes these tubes very attractive for industrial applications.

THE USE of thoriated tungsten instead of pure tungsten for vacuum-tube filaments is desirable for two main reasons. First, the theoretical emission efficiency of thoriated tungsten, which is approximately 100 milliamperes per watt of heating power, is 10 to 15 times that of pure tungsten. Second, the normal range of operating temperatures of thoriated-tungsten filaments is on the order of 1,950 to 2,000 degrees Kelvin, which is 500 to 600 degrees lower than that of pure-tungsten filaments. These features provide definite advantages in the design and application of high-power industrial tubes.

Basic Process

Thoriated-tungsten filament wire is made by adding a small percentage of thoria (thorium oxide) to pure tungsten during manufacture. The thoria content of wire in general use today ranges between 1 and 2 per cent by weight. The procedure required to achieve the high electron-emission capabilities of thoriated tungsten was described by Langmuir[1] in 1923 and has remained substantially unchanged ever since. The filament is flashed at approximately 2,800 degrees Kelvin for a short period (less than 1 minute) to reduce some of the thoria to thorium, and then is operated at a temperature of 2,200 degrees Kelvin or less for a longer period (10 to 15 minutes) to permit diffusion of sufficient thorium to form a monatomic layer on the surface of the wire. Subsequent operating temperatures must be suitably chosen so that a balance is maintained between the production of thorium atoms and their loss from the surface due to evaporation and to positive-ion bombardment.

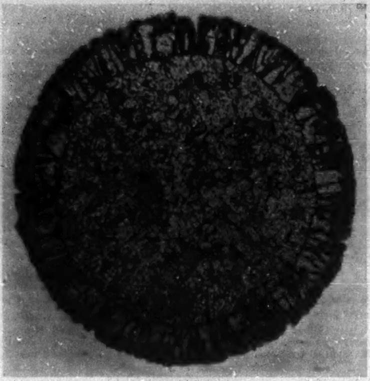
Early Use

Thoriated-tungsten wire was developed originally for use in the lamp industry to prevent off-setting of filaments operating on alternating current. Its electron-emission history dates back to 1913, when it was observed that after suitable heat treatment the emission from thoriated tungsten was considerably higher than that from pure tungsten. Warner and Pike[2] described the use of thoriated-tungsten filaments in low-power transmitting tubes in 1925. Since then, many

---

Paper 53-173, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted June 2, 1952; made available for printing December 23, 1952.

R. B. Ayer is with the Radio Corporation of America, Lancaster, Pa.



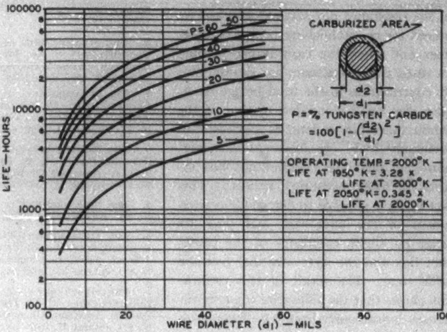

100000
10000
1000
100
8
6
4
2
LIFE—HOURS
0
20
40
60
80
100
WIRE DIAMETER (d_1)—MILS
P=60
50
40
30
20
10
5
CARBURIZED AREA
d_2
d_1
P=% TUNGSTEN CARBIDE
$=100[1-(\frac{d_2}{d_1})^2]$
OPERATING TEMP=2000°K
LIFE AT 1950°K=3.28 x LIFE AT 2000°K
LIFE AT 2050°K=0.345 x LIFE AT 2000°K

Figure 1 (left). Cross section of carburized filament

Figure 2 (above). Thoriated-tungsten filament life curves

other low-power tube types using these filaments have been introduced for radio and industrial use, and a long history of reliable performance has been established. From 1925 to 1945, many attempts were made to use thoriated-tungsten filaments in high-power tubes (those having unmodulated class *C* plate voltage ratings of greater than 5,000 volts), but no success was reported until about 1945.[3] As a result, the general belief developed that these filaments could not be used in tubes operating above 5,000 volts because of rapid deactivation due mainly to high-speed ion bombardment. Statements throughout the literature and in recent books on vacuum tubes[4,5] indicate that this mistaken belief is still widely accepted.

During World War II, however, many radar types were made with thoriated-tungsten filaments. These tubes operated at plate voltages as high as 30 kilovolts and gave very creditable performance.[6–10] Applications in the X-ray field where tubes operated at 100 to 125 kv have also been described recently.[11,12]

Use in High-Power Tubes

Early in 1940, a critical review of the theory and design of thoriated-tungsten filaments was made. At that time, it was concluded that there was no fundamental reason why these filaments could not be used in high-power tubes, provided the mechanical strength and processing factors were given appropriate consideration. Exploratory test models were made of a 20-kw tube type, the *RCA-207*, in which the only deviations from the original structure were the substitution of thoriated tungsten in the filament and the addition of some batalum-type getters. Two of these experimental tubes operated at a plate voltage of 10 kv for 7,953 hours in a 10- to 18-megacycle telegraph transmitter before failure of one tube due to low emission. The grid currents of these tubes, however, were very sensitive to filament input because no precautions had been taken to prevent grid emission. Best performance was obtained with the heating power reduced to 45 per cent of the power required for a corresponding tungsten filament.

Because of the success of the 20-kw type, plans were made to build a 100-kw tube having thoriated-tungsten filaments. Proven mechanical structures were incorporated in this tube, and attention was given to experience gained with low-power thoriated-filament tubes and to the results of numerous filament-emission life tests made at various temperatures on 50-watt types having filaments 8.5 mils in diameter.

Filament Processing

It is customary to carburize a thoriated-tungsten filament to reduce the rate of thorium evaporation from the surface and the susceptibility to deactivation by ion bombardment,[13,14] and thus to achieve longer life. This carburization is accomplished during manufacture by heating the filament to a high temperature in an atmosphere containing a hydrocarbon gas or vapor, as described by Horsting.[15] The result of this process is the formation of a shell of tungsten carbide around the wire as shown in Figure 1. The percentage of the cross section of the wire converted to tungsten carbide is called the "per cent carbide." During life, the filament is gradually decarburized as a result of chemical and thermal processes. Experience has indicated that when substantially all of the carbide is lost, the monatomic layer of thorium can no longer be maintained and the emission decreases. However, the use of a high percentage of carbide, although desirable to assure long life, makes the filament very brittle. A compromise must be made, therefore, between wire size and per cent carbide. Recently, the author presented some curves for use in predicting the emission life of thoriated-tungsten filaments containing various percentages of carbide.[16] Data for these curves were obtained from the life tests referred to previously and from other experience. These curves are shown in Figure 2 in a somewhat more convenient form than in the original presentation. Surprisingly good results have been obtained with these filaments on a wide range of tube types and sizes in spite of differences in construction, processing, and operation. These results represent a useful addition to the available design information on this type of filament.[17–19]

Filament Design

Final designs of a 100-kw tube having a thoriated-tungsten filament showed a reduction in filament heating power of 60 to 70 per cent over pure-tungsten types having equivalent electrical ratings. An example of a water-cooled tube of the 100-kw size is shown in Figure 3. This tube, the *RCA-5770*, can deliver a power output of over 100 kw at frequencies as high as 20 megacycles when used in unmodulated class *C* service with a plate current of 8.5 amperes at 17 kv and a driving power of approximately 2,300 watts. The *RCA-5770* incorporates a self-supported filament structure which requires only 3.1 kw of heating power instead of the 8.1 kw required for its pure-tungsten equivalent. This filament structure, shown in the left portion of Figure 4, was chosen because of its simplicity and its reliable performance in tubes manufactured both here and abroad.[20] The filament consists of a multiplicity of wires connected in series parallel so that the heating current flows in opposite directions in adjacent wires. This arrangement effects the best balance of the magnetic forces and also permits the use of wire crossties to hold alternate filament wires in proper mechanical alignment. These crossties are visible in the assemblies shown in Figure 4. No insulators, springs, or sliding members are required.

Tube Design and Processing

Other design features and processing techniques have contributed to the successful performance of high-power tubes having thoriated-tungsten filaments. Grid emission, which was a problem in the initial 20-kw tubes, has been minimized by the use of platinum-clad molybdenum wire for the grid winding. Oxygen-free high-conductivity copper anodes, widely used in the vacuum-tube industry today, have aided materially in maintaining cleanliness and in minimizing the sources of contaminants. Anodes 1/4 inch thick are used to provide greater mechanical strength and better heat distribution. In addition, although similar tubes built without a getter are still in operation after 20,000 to 30,000 hours, a getter consisting of a small sheet of zirconium is incorporated in such tubes as the *RCA-5770*. The zirconium, heated mainly by radiation, assures a high degree of vacuum over a long period of time. The gettering properties of zirconium at elevated temperatures have been described by Raynor.[21]

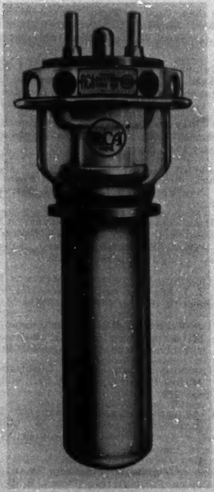


Figure 3. The RCA-5770

A grid support which totally encloses the internal filament-support structure is used in the *RCA-5770* to provide a very-low-inductance grid connection and excellent radio-frequency shielding. This grid support, shown in the center section of Figure 4 also provides good thermal shielding of the glass and minimizes bulb-bombardment effects, thereby reducing the operating temperature of the glass.

A prime requisite in the manufacture of vacuum tubes is cleanliness. In these high-power thoriated-filament tubes, cleanliness is assured by the complete removal of all greases and contaminants resulting from mechanical operations and by a thorough washing and rinsing after chemical or electrochemical cleaning. Oxidation is kept to a minimum during the welding of the final assemblies. The tube elements are outgassed by means of filament radiation and electron bombardment. Gas pressures are measured with an ionization gauge. Pressures are kept low during the initial stages, and processing is continued until a final pressure of 10^{-5} to 10^{-6} millimeter of mercury is reached with the tube hot. Final seal-off is accomplished by squeezing the metal exhaust tube between round, hardened-steel jaws until the tube is severed from the exhaust system.

Final tests made on the *RCA-5770* include measurement of residual gas, of electrical characteristics, and of peak emission. In addition, tubes are operated under dynamic conditions in excess of their ratings. Oscillator tests at 1.5 megacycles have been made at a plate input of 193 kw at 19 kv, which is approximately 24 per cent above its rating. Oscillator tests at 22 megacycles have been made with an input of 153 kw at 17 kv, which is about 25 per cent above the rating of the pure-tungsten-filament type at this frequency. The limiting factor in these oscillator tests was either available rectifier power or failure of circuit components.

Many of the design features and processing techniques used in the 100-kw tubes also were incorporated in tubes of the 50-kw size. A typical water-cooled 50-kw tube is shown in Figure 5. This tube, the *RCA-5771*, can deliver a power output of over 50 kw at frequencies as high as 25 megacycles when used in unmodulated class *C* service with a plate current of 4.5 amperes at 15 kv and a driving power of approximately 1,160 watts. The *RCA-5771* also uses a self-supported thoriated tungsten filament as shown in the right portion of Figure 4. Standard tests on this tube are similar to those made on the *RCA-5770*; special tests have been made at 1.5 megacycles with an input of 90 kw at 15 kv.

A radical departure from conventional high-power tube design was made in constructing a water-cooled super-power tube of the 500-kw size, the *RCA-5831*. This tube, shown in Figure 6, can deliver a power output of approximately 500 kw at frequencies up through the standard broadcast band when used in unmodulated class *C* service with a plate current of 39 amperes at 16 kv and a driving power of approximately 900 watts. The *RCA-5831* utilizes electron-optical principles, thereby improving the power gain by a factor of ten or more. The emitter consists of 48 thoriated-tungsten filaments of trapezoidal cross section held in accurate alignment within the beam-former slots by means of tension provided by a spring-loaded pantograph. Grid rods of similar cross section are held in position over the beam-former in a similar manner. The plate has external spiral slots and is jacketed to improve the cooling efficiency. Normal tests on this type include operation as an amplifier at a plate input of 1,250 kw at 20 kv; occasional tests

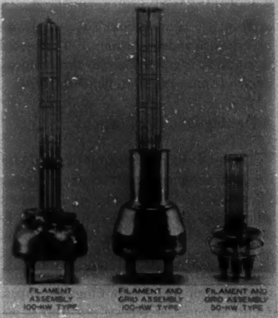

FILAMENT ASSEMBLY 100-KW TYPE
FILAMENT AND GRID ASSEMBLY 100-KW TYPE
FILAMENT AND GRID ASSEMBLY 50-KW TYPE

Figure 4. RCA-5770 filament mount, RCA-5770 filament-grid mount, and RCA-5771 mount

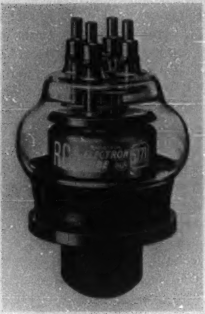


Figure 5. The RCA-5771

have been made at inputs as high as 1,800 kw.

Field Experience

The ability of high-power tubes having thoriated-tungsten filaments to stand up under continuous industrial service has been well demonstrated by field performance. Tubes of the 100-kw size operating at substantially maximum 17-kv ratings in 165-kc tin-reflow oscillators are still in operation after more than 16,000 hours (approximately 2 years) of trouble-free service. Two tubes which had been held as spares for a year and a half were put into service without any sacrifice in performance due to the long period of storage. A 13-megacycle installation operating at 14 kv and about half rated load has given over 22,000 hours (nearly 3 years) of very satisfactory service in a foam-rubber-drying application. In radio broadcast service, the *RCA-5671*, which is similar to the *RCA-5770* with a finned cooler attached, is still in service after 30,000 hours (about $4^1/_2$ years) with no evidence of deterioration. This record is equal in every respect to the performance of pure-tungsten-filament types.

Advantages in Applications

The high-emission capabilities and low-input requirements of thoriated-tungsten filaments result in several important advantages in the application of high-power tubes. Operating costs are reduced as a result of their lower input requirements, which are roughly one third that of equivalent pure-tungsten tubes. As an example, if service per year is assumed to be 7,500 hours (three 8-hour shifts, 6 days per week) and power is assumed to cost 1/2 cent per kilowatt-hour, the savings in the first year more than off-set the differential in the current list price of the pure-tungsten and the thoriated-tungsten 100-kw tube types. In addition, the lower heating-power requirements permit the use of less costly transformers, contactors, and wiring.

The lower operating temperature of the thoriated-tungsten filament results in appreciably lower temperatures of other tube parts. As a result, these parts have a smaller temperature excursion from room to operating conditions, and thermo-mechanical problems are reduced. The reduction in glass and seal temperatures also provides greater safety in high-frequency operation under present ratings, and enhances the possibility of extending the high-frequency ratings. In addition, lower glass and seal temperatures permit some simplification of cooling equipment and may in some cases permit elimination of forced-air envelope cooling.

Even with the reduced heating-power requirements, peak emission two to three times that of the pure-tungsten filament can be provided. The reduction in filament power also makes possible a corresponding increase in anode dissipation. In addition, plate characteristics are less compressed in the high positive grid-voltage region.

Conclusion

Experience to date indicates that the life of the thoriated-tungsten filament can easily be made to equal that of the pure-tungsten type. There are some indications that the life of the thoriated-tungsten type may be appreciably better, all other factors being equal. This possibility seems entirely reasonable if appropriate consideration is given to such factors as carbide loss, gas evolution from all tube parts and particularly from the glass envelope, bulb-bombardment effects, and thoria content of the wire. Although operating experience has been obtained only on tubes through the 100-kw size operating at plate voltages up to 17 kv, there appears to be no practical limitation on either the power or the operating plate-voltage level. The successful use of thoriated-tungsten filaments in high-power tubes thus makes possible a broadening of the range of power sizes available for industrial as well as other uses.

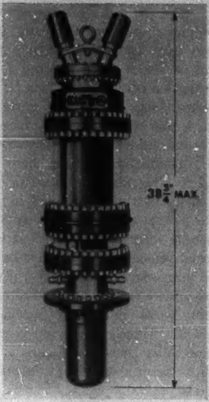


Figure 6. The RCA-5831

References

1. The Electron Emission from Thoriated Tungsten Filaments, **I. Langmuir.** *Physical Review* (New York, N. Y.), volume 22, number 4, October 1923, pages 357–98.

2. The Application of the X-L Filament to Power Tubes, **J. C. Warner, O. W. Pike.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 13, 1925, pages 589–609.

3. Engineering Details of OWI 200 Kilowatt Units, **Hugo Romander.** *Electronics Industries* (New York, N. Y.), volume 4, October 1945, pages 100–03, 158, 162.

4. Vacuum Tubes (book), **K. R. Spangenburg.** McGraw-Hill Book Company, Inc., New York, N. Y., 1948, page 36.

5. Electronics (book), **Jacob Millman, Samuel Seeley.** McGraw-Hill Book Company, Inc., New York, N. Y., second edition, 1951, page 176.

6. A Vacuum-Contained Push Pull Triode Transmitter, **H. A. Zahl, J. E. Gorham, Glenn F. Rouse.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 34, February 1946, pages 66W–69W.

7. Radar Vacuum Tube Development, **J. J. Glauber.** *Electrical Communication* (New York, N. Y.), volume 23, September 1946, pages 306–19.

8. Triodes for Very Short Waves, **J. Bell, M. R. Gavin, E. G. James, G. A. Warren.** *Journal*, Institution of Electrical Engineers (London, England), volume 93, part IIIA, 1946, pages 833–46.

9. A Quarter Century of Electronics, **I. E. Mouromtseff.** *Electrical Engineering*, volume 66, February 1947, pages 171–77.

10. Electron Tubes in World War II, **J. E. Gorham.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 35, March 1947, pages 295–301.

11. High-Voltage Rectifier Valves for X-Ray Diagnostics, **J. H. van der Tuuk.** *Philips Technical Review* (Eindhoven, Holland), volume 8, number 7, July 1946, pages 199–205.

12. Thoriated Tungsten Filaments in Rectifiers, **Z. J. Atlee.** *Electrical Engineering*, volume 68, number 10, October 1949, page 863.

13. The Evaporation of Thorium from Tungsten, **M. R. Andrews.** *Physical Review* (New York, N. Y.), volume 33, 1929, pages 454–58.

14. Thermionic Emission, **S. Dushman.** *Review of Modern Physics* (Lancaster, Pa.), volume 2, 1930, page 449.

15. Carbide Structures in Carburized Thoriated Tungsten Filaments, **C. W. Horsting.** *Journal of Applied Physics* (New York, N. Y.), volume 18, January 1947, pages 95–102.

16. Use of Thoriated Tungsten Filaments in High-Power Transmitting Tubes, **R. B. Ayer.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 40, May 1952, pages 591–94.

17. Properties of Carbonized Tungsten, **B. T. Barnes.** *Journal of Physical Chemistry* (Baltimore, Md.), volume 33, 1929, pages 688–91.

18. The Characteristics of Tungsten Filaments as a Function of Temperature, **H. A. Jones, I. Langmuir.** *General Electric Review* (Schenectady, N. Y.), volume 30, June 1927, pages 310–19.

19. Designing Thoriated Tungsten Filaments, **H. J. Dailey.** *Electronics* (New York, N. Y.), volume 21, January 1948, pages 107–09.

20. High Power Valves: Construction, Testing and Operation, **J. Bell, J. W. Davies, B. S. Gossling.** *Journal*, Institution of Electrical Engineers (London, England), volume 83, number 500, August 1938, pages 176–207.

21. The Use of Zirmet for Gas Absorption **W. M. Raynor.** *Foote Prints* (Philadelphia, Pa.) volume 18, number 2,1947, pages 22–24.

No Discussion

The Elements of Electrocardiographic Theory

ERNEST FRANK
ASSOCIATE MEMBER AIEE

Synopsis: A survey of some of the more important theoretical ideas of electrocardiography is presented for the consumption of the electrical engineer. The development, importance, and objectives of electrocardiography are discussed briefly, followed by a description of the electrical activity of the human heart and the potentials it produces at the surface of the body. Standard connections used in electrocardiography are described. A simplified theory is developed based upon a homogeneous, spherical conducting model containing a current dipole at its center, and the weakness of the assumptions on which it is based is pointed out. The concept and use of the heart vector is treated and methods of portraying heart-vector functions, similar to impedance functions, are illustrated. A limited description is given of the new field of vectorcardiography.

THE electrocardiograph is a sensitive recording voltmeter used to produce a record of the potential differences that arise between various parts of the body from the beating of the heart. This record, called the electrocardiogram, is frequently used in medical examination. However relatively few patients understand the nature of the electrocardiogram or appreciate its usefulness and limitations. Being intimately associated with human heart action the electrocardiogram naturally has great appeal to all, but because it is basically a record of electrical phenomena it should be particularly fascinating to electrical engineers.

"In spite of changing trends in medical education it remains true that the practitioner of medicine does not possess, nor can it be expected that he should possess, a background of training in mathematics adequate to enable him to follow the theoretical developments in electrocardiography."[1] It is hoped that this intriguing subject will arouse the interest of electrical engineers to further investigation of this field where there is need for contributions to the basic theory of the electrical phenomena associated with the human heart beat as well as for new ideas in instrumentation.

Some of the essential ideas related to the electrocardiogram are presented here, and a simplified (but often used) theory of electrocardiography is developed. The presentation is distilled from a wide variety of periodical literature and medical books and is translated into language familiar and digestible to the engineer. Certain liberties have been taken with traditional conventions and nomenclature used in this field in order to simplify the subject. Medical terminology is avoided as much as possible.

The seeds of the field of electrocardiography[2] were planted in 1791 by Luigi Galvani who observed a correlation between electrical excitation and muscular contraction in frogs. Some time passed before Kölliker and Müller discovered (1856) that a frog's heart produced an electric current accompanying each beat. Later in 1878 Sanderson and Page used the capillary electrometer[3] to record the electric currents generated by the exposed heart of a frog. Waller demonstrated in 1887 that the electrical effects of the heart beat could be measured without exposing the heart by using electrodes connected to the body. He was the first to obtain a human electrocardiogram, also using the capillary electrometer. The "father of electrocardiography," William Einthoven, perfected and used in 1903 a more sensitive instrument, the string galvanometer,[3] to record the electrical manifestations of human heart action. He also established[4] certain principles and theories of electrocardiography which are still in widespread use today.

Since the turn of the century it has been recognized that the electrocardiogram could be valuable in obtaining information about one of the most important human organs quickly, conveniently, and without need of gaining direct access to the heart. There has been tremendous activity in this field during the past 50 years. Hundreds of thousands of human electrical records have been taken and interpreted and thousands of articles have appeared in the literature. Improved equipment using vacuum-tube circuits and oscilloscopes has been introduced, and standard commercial recording equipment has been developed. A small army, mostly doctors but also some physicists and engineers, on all continents has contributed continuously to an understanding of the electrical aspects of

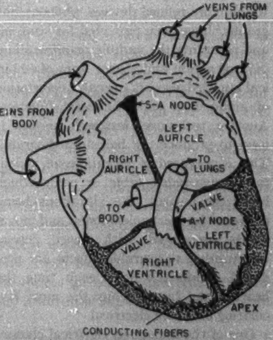


Figure 1. Schematic cutaway front view of the human heart. The apex of the heart is directed downward and toward the left in a normal subject

---

Paper 53-136, recommended by the AIEE Electrical Techniques in Medicine and Biology Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted October 10, 1952; made available for printing December 5, 1952.

Ernest Frank is with the Moore School of Electrical Engineering, University of Pennsylvania, and the Edward B. Robinette Foundation Medical Clinic, Hospital of the University of Pennsylvania, Philadelphia, Pa.

The author acknowledges the many helpful suggestions and constant inspiration offered by Dr. Calvin F. Kay and Dr. S. Reid Warren, Jr., during the pursuit of studies in this field. This work was supported in part by grants of the National Heart Institute, United States Public Health Service. ■

human heart action and the way in which these electrical effects can be used in diagnosis and management of patients with heart disease.

Despite the wealth of knowledge acquired and the huge effort that has been expended, the electrocardiogram is far from completely understood. "It is only one of the objective approaches to an evaluation of the cardiac state and is just as subject to personal bias as any other part of the clinical diagnosis . . . . The interpretation of the electrocardiogram is as subjective as any other form of clinical interpretation, and its value depends . . . on the ability and experience of the electrocardiographer . . . "[5] As the frontiers of knowledge in this field are extended, perhaps this interpretation can become more objective and precise.

Electrical Activity of the Heart

The basic function of the heart is to pump blood through the circulatory system and lungs. The heart consists of two pumping compartments, as indicated schematically in Figure 1, the left side driving the circulatory system of the body and the right side passing blood to the lungs. Blood flowing from the auricles through valves into the ventricles is pumped out with a violent twist-pump action produced by muscular contraction of the ventricles which virtually wrings the blood out into the blood stream.

In the process of pumping, the heart reveals its operations in many ways, each of which has been exploited. One by-product of heart action is acoustic noise, and the simple stethoscope is a well-known device used to listen to the sounds which accompany pumping action. Much more refined devices, phonocardiographs,[6] have been perfected for studying and recording the audible manifestations of heart action and a vast store of empirical knowledge exists to enable correlation of audio data with the heart function. Another technique, less well known, is used to produce the ballistocardiogram.[7] This is a record of vibrations of a delicately suspended table or chair which holds the patient, the vibrations being produced by mechanical forces and movements associated with heart action. Many other measurable effects are produced within the body each time the heart beats, but perhaps the most fascinating of all are electrical.

One of the important electrical characteristics of muscle, of which the heart is almost entirely composed, is polarizability. When heart muscle is in the resting state (not contracted) a potential difference exists between one side and the other of its cellular membranes. This potential difference, which is on the order of 100 millivolts, is explained in terms of modern membrane theory[8] by considering that each element of resting muscle (which is immersed in an electrolyte) is surrounded by a semipermeable membrane through which positive ions may pass freely but which is impervious to the passage of negative ions. A double layer of charges is formed at the membrane, as a result, and constitutes the polarized state which is the condition of heart muscle before it begins its pumping operation.

The heart beat is initiated in the sino-auricular (s-a) node, the pacemaker of the normal heart, indicated in Figure 1. The electrical disturbance periodically created there has been likened to the action of the relaxation oscillator or multivibrator running at a frequency of about 1 cycle per second. The electrical disturbance travels at an average speed of 1 meter per second through the walls of the auricles and causes depolarization of the polarized auricle membranes; that is, it temporarily disrupts the impermeability of the muscle membrane to negative ions whereupon they diffuse outward causing the double layer of charges and the accompanying potential difference to be nullified. In a sense the membrane acts like a battery that is suddenly short-circuited.

The electrical disturbance reaches a second excitation point, called the auriculo-ventricular (a-v) node, at the base of the ventricles where it is delayed by about 0.15 second. Then the impulse spreads rapidly down the wall between the ventricles and along the inner surface of the ventricles through specialized, highly-conducting heart-muscle fibers indicated schematically in Figure 1. This electrical impulse causes the polarized walls of the ventricles to depolarize, the process taking place in the form of a wave of depolarization that spreads outward through the ventricle walls as suggested in Figure 2. The moving double layer, which passes through the right ventricle wall first (which is thinner than the left wall), disappears in about 0.08 second and initiates the mechanical contraction of the ventricles which forces out the blood. The ventricular depolarization wave can be seen to be intimately associated with the pumping function of the heart and is, therefore, of considerable interest in electrocardiography.

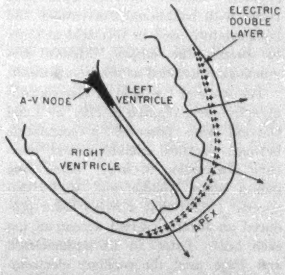


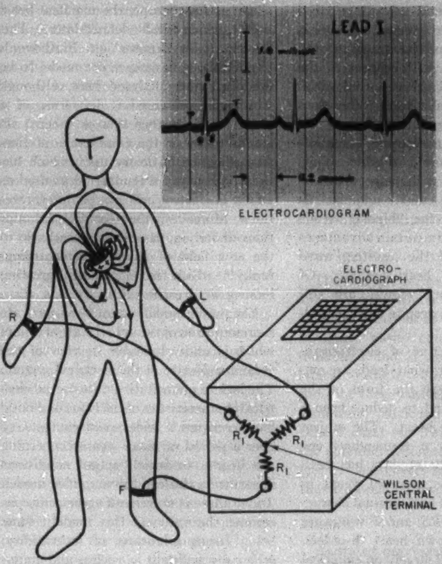
Figure 2. Schematic representation of the electric double layer which moves outward through the ventricle walls initiating muscular contraction of the ventricles

Viewed in gross fashion, the impulse originating in the s-a node can be thought of as quickly depolarizing the polarized resting heart and causing a violent contraction of this intricate blood pump. Another measurable electrical effect occurs approximately 0.37 second after the heart has become completely depolarized when its muscle membranes recover their semipermeable properties thus causing repolarization in preparation for the next impulse from the s-a node.

Potentials at the Body Surface

The sudden changes in potentials across the heart muscles produce electric currents within the body, which is a heterogeneous conductor with resistivity of the order of 10^2-10^3 ohm-centimeters.[9] These currents are distributed throughout the torso but do not penetrate the arms, legs, and neck to an appreciable extent. Body tissues, fluids, and muscle are relatively good conductors; the lungs, which contain air, have about four times the resistivity of tissue; and bones are relatively poor conductors.

A rough suggestion of the current field during the depolarization of the ventricles which are shown as a dome-shaped double layer is given in Figure 3. The strength and shape of the double layer change with time during depolarization and cause the current field to vary. The major frequencies of these currents are in the range 50 cycles per second or less, although some frequency components exist up to about 300 cycles per second.[10] At present, these higher frequency components have no known clinical significance. Although equipment commonly employs bandwidths of 50–100 cycles per second, it is hoped by some electrocardiographers that the use of wider band equipment will enable additional clinical information to be obtained from the electrocardiogram. This represents one of the frontiers of present-day research called high-fidelity



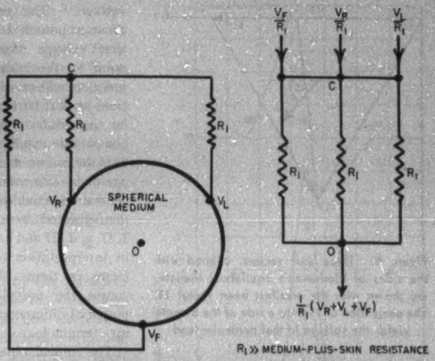


Figure 3 (left). Standard electrodes R, L, and F connected from a human subject to an electrocardiograph. A typical lead I recording during three heart beats is given in the upper right through the courtesy of Dr. C. F. Kay, Hospital of the University of Pennsylvania. The Wilson central-terminal connection is also shown

Figure 5 (above). Wilson central-terminal connection to simplified model along with the node equivalent circuit

electrocardiography.[11]

The currents in the body produce variable potential differences between any two points on the body, a record of which constitutes the electrocardiogram. The potential difference between two points on a given arm or leg is negligible in comparison with that between two points on the torso, as implied by the fact that little current penetrates the limbs. This is an important point in determining the placement of electrodes on the body as will be presently discussed. It can be seen that there are two major influences on the potential differences which appear at the surface of the skin. One is the distribution of current generated at the heart, which can be correlated with certain diseases of the heart; the other is the properties of the medium. The essential value of the electrocardiogram resides in the information contained concerning the generated currents. As a consequence the medium effects play the role of disturbances that distort the desired information.

The standard placement of electrodes (established by Einthoven) is indicated in Figure 3; *R*, *L*, and *F* signifying right arm, left arm, and foot (left leg) respectively. The limbs are chosen because they are essentially equipotential surfaces and therefore the placement of electrodes is not critical and can be reproduced readily in a given patient from time to time, as required. The electrodes usually consist of small german-silver plates strapped firmly to the limb after applying a conducting jelly to the skin which insures good electrical contact. Electrode positions other than on the limbs are also used; in particular, electrodes placed at various standard points on the chest[12] have been found to be of clinical value, but a discussion of these is considered to be beyond the scope of this introductory paper.

Limb-lead voltages are defined as potential differences between pairs of standard electrodes. If V_R, V_L, and V_F signify the potentials at *R*, *L*, and *F* respectively, the three limb-lead voltages are defined as

$$V_I = V_L - V_R$$
$$V_{II} = V_F - V_R$$
$$V_{III} = V_F - V_L \tag{1}$$

and are the potential differences recorded by the electrocardiograph. For brevity, these are commonly referred to as leads I, II, and III. Obviously, these three voltages are not independent being related by

$$V_I + V_{III} = V_{II} \tag{2}$$

which is known in electrocardiography as Einthoven's law although it is nothing more than a statement of Kirchhoff's (1824–87) voltage law. The reason for defining leads I, II, and III with this seemingly curious polarity convention (which is not symmetrical as would be the case defining V_{II} as $V_R - V_F$) stems from the fact that normal deflections produced during ventricular depolarization are such that the dominant swing in voltage is positive with these definitions. This arises basically from the fact that the apex of the heart is directed downward and toward the left as indicated in Figure 3.

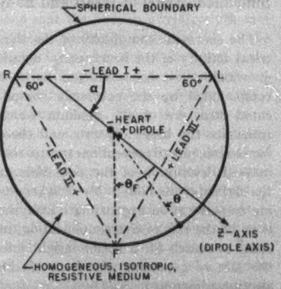


Figure 4. Simplified model of the heart and body which is the basis for a formulation of electrocardiographic theory

A typical electrocardiogram for lead I is shown in Figure 3 along with time and voltage scales. A record of only one lead at a time may be made, or multichannel equipment is available for simul-

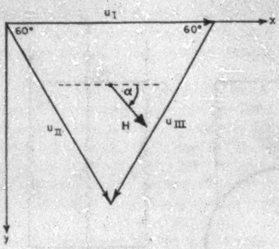


Figure 6. Three unit vectors, aligned with the sides of Einthoven's equilateral triangle, are shown with the manifest heart vector H. The projection of H onto a side of the triangle yields the voltage in that particular lead

taneous lead recording. The record is seen to consist of three separate disturbances called the *P*, *QRS*, and *T* waves, in that order. The *P*-wave results from depolarization of the auricles, the *QRS*-wave (or *QRS*-complex as it is more frequently called) arises from ventricular depolarization, and the *T*-wave is produced by repolarization of the heart muscle. Triple symbols are used for the ventricular wave, which is much larger in amplitude than the *P* or *T* waves, in order to describe the pulse in more detail; *Q* signifies the first downward deflection, *R* the positive peak, and *S* the final downward overshoot. In more complicated cases of multiple irregularities in pulse shape, subscripts are employed on *Q*, *R*, and *S* to distinguish among various upward and downward deflections. It is informative to note that the maximum lead voltage is about 1 millivolt which is only a small fraction of the potential difference across the double layer in the heart. This comes about because the current density diminishes markedly with distance from the heart; consequently, most of the potential drop occurs within the body.

An external connection called the Wilson central terminal[13] is commonly used. It consists of three equal resistors R_1 which are y-connected from *R*, *L*, and *F* to a common junction as shown in Figure 3 and has been the subject of wide controversy and much experimentation. It is claimed that the potential of this junction (with respect to the potential at infinity) does not vary appreciably with time; the theoretical foundation for this is based upon the idealized model to be described. Some experiments have been performed from which it has been deduced that the Wilson central-terminal potential variations are on the order of 20 per cent or less than the maximum lead voltage.[14] The prime advantage of a constant-potential reference junction is that voltage measurements from any point on the body with respect to the junction will reveal the potential variations at that particular point rather than be the result of a combination of two comparable-magnitude varying potentials as is the case using standard leads. Voltage measurements with respect to the central terminal are called "unipolar" as distinguished from the "bipolar" leads I, II, and III and have certain advantages in interpretation of the resulting wave forms in terms of heart activity. Of course the unipolar voltages are still potential differences despite this misleading terminology.

The clinical objective of electrocardiography, no matter what leads are employed, is to analyze the form of the electrocardiogram and to deduce from it the condition of the heart. The subject of interpretation is a tremendous one and a huge quantity of data has been assembled over the past 50 years in which certain quirks and unusual characteristics of the *P*, *QRS*, and *T* waves are associated with known heart disorders. These data are based largely on empirical information, such as post-mortem anatomical findings. Electrocardiographic theory has been very useful in providing a mathematical basis for analysis of the electrocardiogram itself, but has more limited value in interpretation which is still strongly dependent upon empiricism. It is the hope of theoretical workers in this field that continued refinements of the theory will progressively reduce the element of empiricism and supplant it by theoretical considerations which will give greater insight into the form of the electrocardiogram and thereby increase its usefulness.

Simplified Model of Heart and Body

The enormous complexity of the electrical activity of the heart, many details of which have been omitted here, is compounded by the extremely complicated structure of the medium which transmits the heart currents and these factors conspire to prevent an exact mathematical description of the potentials at the surface of the body. Yet it is tempting to perform an approximate analysis, even at the risk of oversimplification, in order to search for a better insight into the factors governing the form of the electrocardiogram.

In 1913 Einthoven[4] described what he called a "schema" that he had found useful in consolidating limb-head data into a single useful concept, the manifest heart vector, which will be defined later. This schema, now known as Einthoven's hypothesis, is in essence the model to be described and analyzed here, although Einthoven presented it in terms of a plane lamina rather than a sphere. It stands today as the cornerstone of electrocardiographic theory upon which has been built many secondary ideas used as an aid in electrocardiographic interpretation. Moreover, the essential assumptions underlying the schema are basic to the new field of spatial vectorcardiography[15] which is in rapidly growing infancy at present.

The simple model cannot be an accurate representation of the actual state of affairs which is understandable in view of the great complexity of the electrical system. The lack of quantitative data on the electrical characteristics of the heart and body has permitted a widespread controversy over a period of many years concerning the degree to which actual conditions depart from those of the simplified model. Today there is no general agreement concerning the value of this model; some being strong advocates of its validity, others discarding it as useless and claiming that it has stunted the growth of the field.

The following assumptions concerning the heart and body define the simplified model that is widely used today:

1. Boundary. It is assumed that the resistivity of air is much greater than that of the body and that the surface of the body can be represented by a sphere.
2. Limb Leads. It is assumed that the limb electrodes are located electrically on the sphere at the corners of an equilateral triangle lying in an equatorial plane of the sphere, this plane being called the frontal plane.
3. Medium. It is assumed that the body is a homogeneous, isotropic, resistive medium.
4. Heart. It is assumed that the heart can be represented by a current dipole of variable strength and orientation but fixed in position at the center of the spherical conducting medium.

These assumptions are represented in Figure 4. The dipole axis, a line joining the current sources, is shown to intersect lead I at an angle α, measured positive in the clockwise direction which is the convention established by Einthoven and firmly rooted today in all of the literature. Using standard spherical co-ordinates (taking the dipole axis to be the *z*-axis) the polar angle between the dipole axis and the radial line to *F* is $\theta_F = 90° - \alpha$. The dipole shown in Figure 4 is really the projection of the actual heart dipole onto the frontal plane and is more accurately termed the effective heart dipole. The actual heart dipole can be resolved into two dipoles at right angles to one another; the one in the frontal plane being shown in Figure 4 while the one perpendicular to the frontal plane is ignored since it does not contribute to the potentials at R, L, and F when the position of the dipole is at the sphere center. It follows that the z-axis of Figure 4 is the projection of the actual dipole axis onto the frontal plane.

The principal advantage of the assumptions is that they result in an extremely simple model that can be analyzed easily and which has some consequences of considerable mathematical nicety. However, it is at best a gross approximation to actual conditions. The philosophy underlying the use of such a model, despite its limitations, has been expressed by Dr. F. N. Wilson, a leading authority on electrocardiographic theory.

"... the vast majority of those who have been engaged in the study of the human electrocardiogram have had small acquaintance with, and little interest in, mathematical attacks upon physical problems of the sort encountered in attempts to apply the classical theory of electricity to the analysis of the varying electrical field associated with the heart beat. In theoretical investigations of this kind the actual situation under consideration is always far more complicated than any of those that can be treated mathematically, and it is necessary to make simplifying assumptions that are not strictly in accord with the facts. To assert that all deductions based on such assumptions are *ipso facto* worthless is, so to speak, to deny that mathematics has contributed anything worthwhile to the physical sciences. To maintain, on the other hand, that deductions of this kind represent anything more than a first approximation to the truth or have any great value except in so far as they are supported by experience and by experiments designed to test their validity would be equally unreasonable. It is imperative that those who make use of conclusions of this sort as a guide to further investigations, or who attempt to extend them, clearly understand and constantly bear in mind the postulates upon which they rest."[16]

Heeding the warning in the last sentence, it is fitting to look into some of the more glaring defects of the assumptions.

BOUNDARY

The most valid assumption of all is that the resistivity of air is much greater than that of the body, which averages about 500 ohm-centimeters. The assumption of a spherical shape is not as grossly wrong as might appear from purely geometrical considerations. Since the current generated by the heart penetrate only slightly into the shoulders, thighs, and neck there is small error in discarding the arms, legs, and head. Undoubtedly an ovoid or ellipsoidal boundary would improve the accuracy of the analysis, but a spherical boundary has the great advantage of simplicity.

LIMB LEADS

The triangle formed by the shoulders and left hip (which are the equivalent electrical locations of the limb electrodes) obviously varies considerably with the individual and is not exactly an equilateral triangle. However, the importance of these variations may not be too significant particularly in view of the experimental fact that in most patients the potential difference between the right and left legs is very small compared with that between leg and arm.

MEDIUM

As stated previously, the body is comprised of substances with widely differing resistance. Only limited quantitative data are available concerning the resistivity of various constituents of the body, and moreover, the effect of these inhomogeneities on the limb-lead voltages is open to debate. Indeed, a major controversy has been raging for years concerning the degree to which the medium influences the electrocardiogram. On one hand, it is maintained that there are certain highly favored paths for current conduction in the body which are in good contact with the heart, and consequently, that these favored parts of the heart have a dominant influence on the form of the electrocardiogram. On the other hand, many electrocardiographers contend that the body is a relatively homogeneous conductor and that all of the double-layer elements of the heart make contributions to the potentials at the surface of the body. Corresponding to these divergent views, there is a cleavage in thought on interpretation of the electrocardiogram with the former school of thought adhering to the idea that it be used almost entirely as an empirical tool and the latter groups placing emphasis on simplified homogeneous volume conductor theory of the kind to be presented here. Of course there are many who straddle both views and this seems to be a sensible choice until more knowledge is acquired.

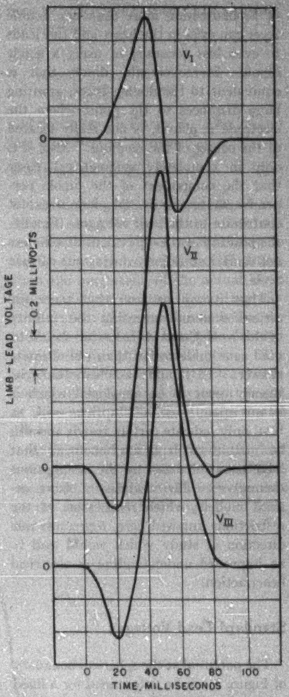


Figure 7. Typical limb-lead voltages during the QRS-complex (adapted from Einthoven)

The medium presents additional difficulties. There is evidence that muscle fibers display certain favored directions for current flow, which casts doubt on the isotropic assumption. In addition, polarization phenomena, distinctly different from heart-muscle polarization previously mentioned, is known to be important in physiological substances and may play some role in influencing the voltage distribution. There is much more to be learned about the electrical characteristics of the medium before any final answer to the medium assumption can be offered.

HEART

The double layer which produces the *QRS*-complex can be represented by an equivalent dipole if the distance from the double layer is very large compared with the dimensions of the double layer.[17] Since the chest diameter is only about three times the diameter of the ventricles the distance requirement is not completely fulfilled, but the error of the approximation has not been evaluated quantitatively. In addition, during some instants of depolarization more than one double layer can exist in the heart and this leads to even less accuracy in using a single dipole. Moreover, the dipole that is equivalent to the double layer, granting large distances to the point where the electrode is placed, is not really located at the center of the medium. Also, it is only in completely symmetrical cases that the component of the dipole perpendicular to the frontal plane does not contribute to the lead voltages. Finally, the position of the equivalent dipole does not remain stationary during the cardiac cycle.

Thus it can be seen that there are serious doubts concerning the validity of this model, and its behavior should be used as a guide rather than the ultimate answer. It should be clear that the theory about to be developed which is based upon these assumptions will, at best, only indicate certain trends and will be unable to explain phenomena that have been ruled out by the assumptions themselves. The analysis of more refined models, which relax some of the restrictions imposed here, represents one direction of study which would lead to an improved understanding of electrical heart action.

Standard Lead Voltages

The potential at the spherical boundary of Figure 4 can be determined for a fixed orientation of the effective current dipole and the results of such an analysis can be used for cases in which the effective dipole strength and orientation are functions of time since the medium is assumed to be resistive and the frequencies of the currents are very low. In order to determine the potential at the boundary of the sphere of radius R in the d-c case, it is necessary to solve Laplace's equation subject to the condition that the normal derivative of the potential is zero at the boundary (which is equivalent to the requirement that no current lines shall pierce the boundary). The potential at any point in the medium a distance $r<R$ from the dipole is[18]

$$V_r=\frac{p}{4\pi\gamma}\left(\frac{1}{r^2}+\frac{2r}{R^3}\right)\cos\theta \qquad (3)$$

where p is the dipole moment, γ is the conductivity of the medium, and θ is the standard polar angle shown in Figure 4. The solution satisfies Laplace's equation as well as the boundary condition

$$\left.\frac{\partial V}{\partial r}\right|_{r=R}=0$$

and also reduces to a dipole field in an infinite medium if R is so large in comparison with r that the second term of the sum is negligible in comparison with the first term.

Since the electrocardiographer is concerned with potentials at the boundary, Equation 3 can be reduced to this case by imposing $r=R$.

$$V=\frac{3p\cos\theta}{4\pi\gamma R^2} \qquad (4)$$

Expressions for the standard limb-lead voltages can be obtained by recognizing that F, R, and L are defined by the angles $\theta=\theta_F$, $\theta_F+120°$ and $\theta_F-120°$ respectively. Furthermore, with $\theta_F+\alpha=90°$ it follows from equation 4 that

$$\begin{aligned} V_R&=\frac{3p}{4\pi\gamma R^2}\cos(\theta_F+120°)\\ &=\frac{3p}{4\pi\gamma R^2}\cos(210°-\alpha)\\ V_L&=\frac{3p}{4\pi\gamma R^2}\cos(\theta_F-120°)\\ &=\frac{3p}{4\pi\gamma R^2}\cos(-30°-\alpha) \qquad (5)\\ V_F&=\frac{3p}{4\pi\gamma R^2}\cos\theta_F=\frac{3p}{4\pi\gamma R^2}\cos(90°-\alpha) \end{aligned}$$

Using the definitions of the limb-lead voltages, equation 1, straightforward trigonometric manipulation of equation 5 leads to

$$\begin{aligned} V_{\rm I}&=\frac{3\sqrt{3}p}{4\pi\gamma R^2}\cos\alpha\\ V_{\rm II}&=\frac{3\sqrt{3}p}{4\pi\gamma R^2}\cos(\alpha-60°) \qquad (6)\\ V_{\rm III}&=\frac{3\sqrt{3}p}{4\pi\gamma R^2}\cos(\alpha-120°) \end{aligned}$$

which are the desired limb-lead voltage equations.

The expressions in equations 5 and 6 have the same form as those which pertain to a balanced 3-phase y-connected power system with V_R, V_L, and V_F resembling phase voltages and $V_{\rm I}$, $V_{\rm II}$, and $V_{\rm III}$ resembling line voltages (note the $\sqrt{3}$ factor) except the reversed polarity convention used in electrocardiography for $V_{\rm II}$ destroys the symmetry. This similarity in form may be taken advantage of in deriving equation 6 directly from a phasor diagram of V_R, V_L, and V_F. Of course Einthoven's law, equation 2, is satisfied by the lead voltages in equation 6 for any values of p and α, which may be considered to be slowly varying functions of time.

It can be seen from equation 6 that variations in the limb-lead voltages are completely specified by two properties of the effective heart dipole, its moment p and its orientation in terms of the angle α.

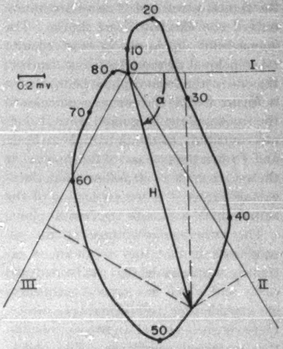


Figure 8. Monocardiogram constructed from the limb-lead voltages of Figure 7. The points on the locus of the tip of H indicate the time in milliseconds. The dashed lines indicate the projections of H onto the three limb-lead directions

The simplicity of the model has led to the conclusion that the variation in lead voltages depends only upon these two important properties of the dipole for fixed body geometry. It has been shown recently[19] that for a completely general shaped boundary and for an inhomogeneous medium that a similar type of conclusion can be drawn, provided the assumption of a fixed position dipole representation of the heart is retained. However, the component of the heart dipole that is perpendicular to the frontal plane does contribute to some extent to the limb voltages in the general case. Thus, the simplicity of the model, which has a high degree of symmetry, has led to an erroneous idea regarding details but nevertheless has led to the same type of conclusion reached in a less restricted analysis.

Wilson Central Terminal

The expressions for the limb-lead voltages can be used to show that the potential of the Wilson central terminal[13] is constant. The most expedient way to do this is in terms of node equations which in this special case can be regarded as an application of Millman's theorem.[20] The central terminal arrangement is shown in Figure 5 along with the node equivalent circuit in which it has been assumed that the y-resistors R_1 are very large in comparison with the internal resistance behind the electrode voltages. The point 0 signifies the centerpotential of the dipole which is arbitrarily assigned the reference value zero. The potential difference between the central terminal C and the center potential of the dipole can be written immediately from Figure 5 as the sum of the currents of the current generators times the total resistance between C and 0.

$$V_{C_0} = \frac{1}{R_1}(V_R + V_L + V_F)\frac{R_1}{3}$$
$$= \frac{V_R + V_L + V_F}{3} = 0 \quad (7)$$

The sum $V_R + V_L + V_F$ is zero for any values of p and α as can be shown by straightforward trigonometry using the expressions in equation 5. This becomes obvious by again considering the phasor representations of V_R, V_L, and V_F since these voltages have equal amplitude factors and are symmetrically oriented at 120° apart. Thus, the theoretical basis has been shown for the claim that V_{C0} will not vary with time despite variations in the strength and orientation of the heart generator.

If the R_1 are not large enough, they will load the model. Originally, Wilson used values of $R_1 = 5{,}000$ ohms, but in current practice values of 25,000 to 50,000 ohms are used, the upper limit on R_1 being imposed by practical difficulties with stray pickup. The fact that the Wilson central-terminal voltage is not exactly constant in practice is not due to the loading effects of R_1 but rather to the defects in the assumptions which have led to the zero potential conclusion. Indeed, it can be shown on the basis of the simplified model but using a dipole that is not located at the center of the sphere that V_{C0} is not zero, even theoretically.[21]

Heart-Vector Theory

The great simplicity of equation 6 enables the limb-lead voltages to be represented geometrically, the phasor representation being one that has been mentioned already. A similar representation can be formulated in terms of vectors which enables each lead voltage to be considered as arising from a dot product of two vectors. This latter idea has been used widely in electrocardiography.

The phasor representation can be established in a manner similar to that used in a-c circuit theory, except the phasors are not considered to rotate at a fixed angular velocity since there is no assurance that α is a constant times the time. Defining $H = 3\sqrt{3}p/4\pi\gamma R^2$, which is a constant times the effective dipole moment p (for fixed radius and conductivity of the model), then the phasors representing the lead voltages can be written as

$$\begin{aligned} V_{\mathrm{I}} &= H\epsilon^{j\alpha} = H\epsilon^{j\alpha}\epsilon^{0} \\ V_{\mathrm{II}} &= H\epsilon^{j(\alpha-60°)} = H\epsilon^{j\alpha}\epsilon^{j-60} \qquad (8) \\ V_{\mathrm{III}} &= H\epsilon^{j(\alpha-120°)} = H\epsilon^{j\alpha}\epsilon^{-j120°} \end{aligned}$$

where $j^2 = -1$. These complex numbers can be interpreted as the product of a heart dipole phasor $H\epsilon^{j\alpha}$ times phasors of unit magnitude directed along leads I, II, and III respectively. The lead voltages are the real parts of these complex numbers.

$$\begin{aligned} V_{\mathrm{I}} &= Re(\mathbf{V}_{\mathrm{I}}) = H\cos\alpha \\ V_{\mathrm{II}} &= Re(\mathbf{V}_{\mathrm{II}}) = H\cos(\alpha - 60°) \qquad (9) \\ V_{\mathrm{III}} &= Re(\mathbf{V}_{\mathrm{III}}) = H\cos(\alpha - 120°) \end{aligned}$$

and can be interpreted geometrically as the projection of the heart phasor onto the unit phasors directed along each lead.

The customary viewpoint taken in electrocardiography is based upon a vector representation. To establish this formally, define two unit vectors $\mathbf{i}$ in the positive direct of lead I and $\mathbf{j}$ perpendicular to $\mathbf{i}$ and directed downward toward F. These are the standard unit vectors of 2-dimensional rectangular co-ordinates. Expressions for three unit vectors of unit magnitude whose directions coincide with the sides of the equilateral triangle may be written in terms of the standard unit vectors

$$\begin{aligned} \mathbf{u}_{\mathrm{I}} &= \mathbf{i} \\ \mathbf{u}_{\mathrm{II}} &= \mathbf{i}\cos 60° + \mathbf{j}\sin 60° \qquad (10) \\ \mathbf{u}_{\mathrm{III}} &= \mathbf{i}\cos 120° + \mathbf{j}\sin 120° \end{aligned}$$

and are shown in Figure 6. The "manifest heart vector" is defined as

$$\mathbf{H} = \mathbf{i}H\cos\alpha + \mathbf{j}H\sin\alpha \qquad (11)$$

which is a vector of magnitude H and directed along the dipole axis. The term manifest was introduced by Einthoven to signify that the magnitude of $\mathbf{H}$ differs from the magnitude of the actual heart dipole moment because of the short-circuiting effect of the medium as well as the fact that the actual heart dipole does not necessarily lie in the frontal plane. The word manifest stems from the idea that $\mathbf{H}$ represents the heart activity as manifested at the surface of the body in the frontal plane.

With these vectors defined, the lead voltages can be considered to arise from vector dot products of $\mathbf{H}$ with the unit vector associated with any given lead. For example

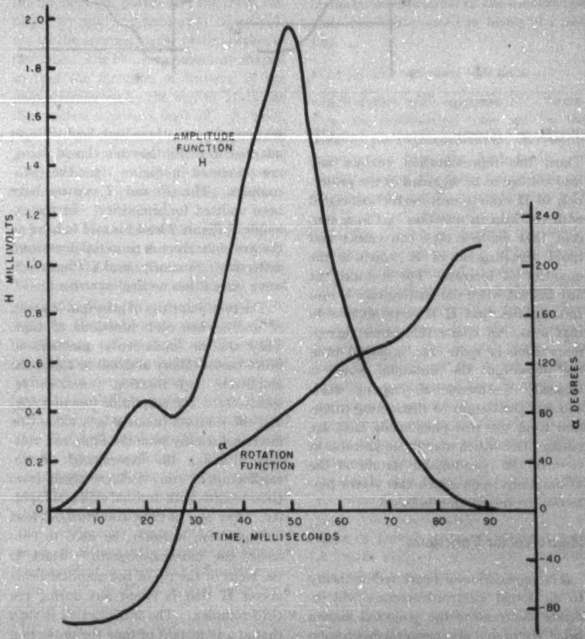
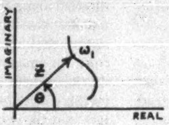


Figure 9. Magnitude and angle of H as a function of time corresponding to the limb-lead voltages of Figure 7 and the monocardiogram of Figure 8

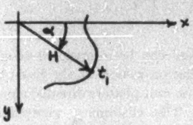
Table I. Correspondence Between Impedance Functions and Heart-Vector Functions

	Impedance Functions	Heart-Vector Functions
Variable	angular frequency ω	time t
Function	complex impedance $\mathbf{Z}=Z\cos\theta+jZ\sin\theta$ $\mathbf{Z}=R+jX$	manifest heart vector $\mathbf{H}=\mathbf{i}H\cos\alpha+\mathbf{j}H\sin\alpha$ $\mathbf{H}=\mathbf{i}H_x+\mathbf{j}H_y$
Magnitude	$Z=(R^2+X^2)^{1/2}$	$H=(H_x^2+H_y^2)^{1/2}$
Angle	phase angle θ $\tan\theta=\frac{X}{R}$	Einthoven's angle α $\tan\alpha=\frac{H_y}{H_x}$

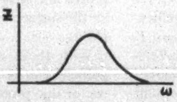
variable-frequency locus in complex plane:



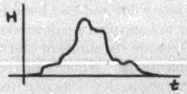
variable time locus in *xy*-plane. Monocardiogram:



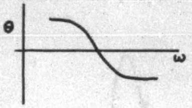
magnitude of **Z** versus ω:



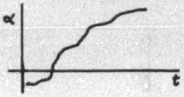
magnitude of **H** versus t. Amplitude function:



θ versus ω:



α versus t. Rotation function:



$$V_{\text{I}}=\mathbf{H}\cdot\mathbf{u}_{\text{I}}=H\cos\alpha \qquad (12)$$

Thus, this representation enables any lead voltage to be regarded as the projection of **H** onto a unit vector associated with the lead in question. It is a concept that enables very convenient and rapid visualization of **H** which is the quantity of interest. For instance, at any instant when the lead voltage is zero this signifies that **H** is perpendicular to that lead. Of course this geometric representation is quite far removed from the physics of the situation, which is basically 3-dimensional current fields and has the danger of distracting attention from the real phenomena that are taking place which may be undesirable in view of the questionable nature of the assumptions upon which this vector-projection representation is based.

Heart-Vector Functions

The application of heart-vector theory to an actual electrocardiogram will indicate the trend of the projected human heart vector and also permit a description of some additional ideas that are important in electrocardiography. For illustrative purposes, three limb-lead voltages adapted from Einthoven's classic paper[4] are presented in Figure 7 for the *QRS*-complex. The *P* and *T* waves have been omitted for simplicity. In the example of Figure 7 lead I is said to have no *Q*-wave since there is no initial downward deflection. Similarly, lead III has no *S*-wave since it has no final overshoot.

The two quantities of principal interest, *H* and α, are each functions of time. They are the fundamental functions of heart-vector theory and will be called the amplitude and rotation functions respectively. The amplitude function $H(t)$ and the rotation function $\alpha(t)$ cannot be discerned readily from the limb-lead voltages although the experienced electrocardiographer can visualize them from these records with unusual skill and rapidity. Any pair of these simultaneous lead voltages can, however, be used to construct the monocardiogram,[22] which is the locus of the tip of the manifest heart vector **H** that is swept out during the *QRS*-complex. The construction is such that at any instant of time the projection of **H** onto any one of the three unit vectors associated with each lead is equal to the instantaneous voltage in that lead. The monocardiogram corresponding to the lead voltages in Figure 7 is given in Figure 8 where various instants of time in milliseconds during *QRS* are indicated by the points marked on the locus. The manifest heart vector is seen to rotate in a clockwise direction, in this particular case, traversing a total angle of approximately 300 degrees and reaching a maximum value of 2 millivolts at t = 50 milliseconds.

Another way in which $H(t)$ and $\alpha(t)$ can be portrayed is given in Figure 9. This alternative way of plotting the data sometimes reveals properties of $H(t)$ and $\alpha(t)$ more readily than in the monocardiogram. In this example, notice how the notch in the amplitude function at t = 25 milliseconds, which is accompanied by a very rapid change in the rotation function, shows up more distinctly than in the monocardiogram of Figure 8.

The major virtue of heart-vector theory is shown by either of these forms of presenting the data. It enables the limb-lead data, which may be difficult to decipher, to be unified into two heart-vector functions which are consistent with the limb-lead data and which consolidate all of the lead information into a coherent whole.

The two methods of presenting heart-vector data are very similar to those used by electrical engineers to portray impedance functions, as might be expected from the close correspondence between the phasor and vector formulations that have been presented. The correspondence between impedance functions and heart-vector functions is given in Table I.

Many aids have been developed to enable the electrocardiographer to determine *H* and α conveniently from limb-lead data. Several different types of charts [22,23] are available which permit *H* and α to be read directly after entering the charts with known simultaneous values of the lead voltages. These charts are similar in nature to impedance charts which can be entered with resistance and reactance values to read off the magnitude and phase of the corresponding impedance. In addition, extensive tables[24] have been constructed for the range of values of V_{I}, V_{II}, and V_{III} encountered in practice which can be used to look up values of *H* and α corresponding to measured values of lead voltages. Also, a special slide rule[25] has been devised to speed the determination of *H* and α from the limb-lead data. Of course, it is essential to have simultaneous lead-voltage data in order to compute the instantaneous values of the heart vector.

The lead voltages given in Figure 7 were obtained by Einthoven with relatively narrow-band equipment (approximately a 50-cycle-per-second bandwidth), and therefore are quite smooth. High frequency wiggles and more abrupt changes are obtained with broad-band equipment and, moreover, certain hearts have much more intricate *QRS*-complexes, sometimes giving rise to multiple-loop monocardiograms and corresponding notches in the rotation function. A more complicated monocardiogram, to be discussed later, is given in Figure 10 to illustrate this point.

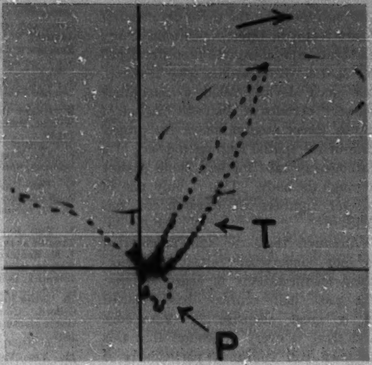


Figure 10. Frontal-plane vectorcardiogram described in the following statement by Dr. W. R. Milnor, Johns Hopkins Hospital, Baltimore, Md., who was kind enough to supply the record: "It was recorded from a 39 year old man who had suffered an acute myocardial infarction (dead region of ventricle wall)." "The P loop is the small loop at about 5 o'clock, and the T loop is long and narrow and points to about 1 o'clock. The rest of the record is QRS, showing a marked slowing in the terminal portion. The timing signals represent 4 millisecond intervals; i.e., 4 milliseconds elapsed from the beginning of one dash to the beginning of the next. The dashes are wedge-shaped and point in the direction of rotation. [illegible] Two grid lines have been drawn in intersecting at the null point"

In clinical application of heart-vector theory, many quantities associated with the amplitude and rotation functions are used in an effort to find a correlation with certain types of heart ailments. For instance, the value of α for which H is a maximum, the area enclosed by the monocardiogram, the total angle through which the vector sweeps, and other such special features are studied in a search for significant characteristics. The treatment of the data is sometimes more elaborate than is perhaps warranted by the questionable model on which heart-vector theory is based. The fact that $H(t)$ and $\alpha(t)$ can always be determined uniquely for a given pair of lead voltages should not delude one into thinking that they are necessarily significant. In truth, an artificial agreement is established by forcing **H** to yield the measured limb-lead voltages, via vector projection. The acid test of the clinical value of the manifest heart vector, so derived, resolves itself into determining the degree to which this vector agrees with the instantaneous distribution and strength of the depolarization wave in the heart muscle. Unfortunately, there is no easy means for applying this test and the correlation between the manifest heart vector and the electrical activity of the heart remains elusive.

Vectorcardiography

It has been shown that the amplitude and angle of the manifest heart vector are the quantities of focal interest to the electrocardiographer; yet these quantities are not directly observable in the electrocardiogram. A relatively new branch of this field, called vectorcardiography, is one in which H and α are obtained directly as a trace on the face of a cathode-ray tube. Hence, the lead voltages assume, appropriately, a secondary role; the manifest heart vector is observed and dealt with directly and the problem of deriving H and α by calculations based upon limb-lead data is circumvented.

When the monocardiogram is displayed directly on an oscilloscope, the record is called a vectorcardiogram or vector loop. Figure 10 shows an example of a vectorcardiogram. A time scale is provided by the blanked-out portions of the trace, obtained by applying accurately timed pulses to the intensity grid of the cathode-ray tube. The blanking pulses are shaped so that the direction of traverse of the loop is indicated by the way in which the illuminated segments trail off. *P*, *QRS*, and *T* loops are swept out each time the heart beats and the lead voltages are similar enough in shape from one beat to the next (see Figure 3) so that a stationary pattern is obtained.

Several methods can be employed in practice to produce the vectorcardiogram, the theory behind these methods being based on the model equations. One method is to apply one of the lead voltages to the horizontal amplifier of the oscilloscope and to apply to the vertical amplifier the voltage of the third electrode (unused in the horizontal signal) with respect to the Wilson central-terminal voltage. For example, V_I may be introduced into the horizontal amplifier and V_F to the vertical amplifier. The voltage amplification of the vertical amplifier must be adjusted to be $\sqrt{3}$ times that of the horizontal amplifier. To show that the oscilloscope trace obtained under these conditions is the same as the locus of the manifest heart vector, consider equation 12, $V_\mathrm{I} = H \cos \alpha$, and equation 5 which becomes

$$\begin{aligned} VF &= (H/\sqrt{3}) \cos (90^\circ - \alpha) \\ &= (H/\sqrt{3}) \sin \alpha \end{aligned}$$

when using $H = 3\sqrt{3}p/4\pi\gamma R_3$. Now, since V_I produces horizontal deflections (x-axis of Figure 6) with which the unit vector **i** has been associated, and $\sqrt{3}V_F$ produces vertical deflections (y-axis of Figure 6) with which the unit vector **j** has been associated, it follows that the instantaneous position of the oscilloscope trace can be expressed in terms of **i** and **j** as

$$\mathbf{i}V_\mathrm{I}+\mathbf{j}\sqrt{3}V_F=\mathbf{i}H\cos\alpha+\mathbf{j}H\sin\alpha$$

which agrees with equation 11. Therefore, the oscilloscope trace will be the same as the monocardiogram derived by construction from the limb leads. In similar fashion, the same vector loop can be produced using V_II in combination with $\sqrt{3}V_L$ or by using V_III with $\sqrt{3}V_R$. It is also possible to produce the vectorcardiogram using any pair of leads without the Wilson central terminal.

The vectorcardiograms discussed here are more accurately called frontal-plane vectorcardiograms since they represent projections on the frontal plane. The locus of the tip of the heart vector is usually a complicated loop in space, not necessarily confined to any plane, and viewing only one projection of this loop limits the amount of information that is obtained. Recognition of this has led to the birth of a new field, called spatial vectorcardiography,[15] in which the frontal-plane vectorcardiogram is only one of three or more projections that are used. A fourth electrode is required to obtain the other projections, and a variety of electrode positions is used, not necessarily the limb electrodes in some cases. The over-all data can be presented as a loop in space, and a wire model of the locus has been found useful for describing the path taken by the tip of the manifest heart vector in three dimensions. Very elaborate equipment has been developed for spatial vectorcardiography research; one such device called a panoramic viewer[26] permits, at the twist of a pair of knobs, direct observation of the projection of the spatial loop on the face of a cathode-ray tube for any plane passing through the body. This is only one of several frontiers of research in this field.

Although there is still much that is not understood about the electrical manifestations of heart activity, the past half century has resulted in a great increase of knowledge and there is every indication that this advance will continue, particularly if electrical engineers can be called upon to help with problems that require more technical knowledge of physical science than is usually found in doctors with medical-school backgrounds.

References

1. An Outline of Electrocardiography, **R. Ashman.** *New International Clinics* (Philadelphia, Pa.), volume 1, 1939, page 193.

2. The History of Electrocardiography, **W. G. Leaman, Jr.** *Annals of Medical History* (New York, N. Y.), volume 8, 1936, page 113.

3. Medical Physics (book), **O. Glasser,** editor. Electrocardiography, **F. D. Johnston.** Year Book Publishers, Inc., Chicago, Ill., 1944.

4. Über die Richtung und die manifeste Grösse der Potentialschwankungen im menschlichen Herzen und über den Einfluss der Herzlage auf die Form des Elektrokardiogramme, **W. Einthoven, G. Fahr, A. deWaart** (In English). *American Heart Journal* (St. Louis, Mo.), volume 40, 1950, page 163.

5. Electrocardiography (book), **L. N. Katz.** Lea and Febiger, Philadelphia, Pa., 1946.

6. Heart (book), **A. A. Luisada.** Williams and Wilkins, Baltimore, Md., chapter 5, second edition, 1948.

7. The Ballistocardiograph, **I. Starr.** Harvey Lectures, Springfield, Ill., 1946–47, page 194.

8. The Genesis of the Electrocardiogram, **L. N. Katz.** *Physiological Reviews* (Washington, D. C.), volume 27, 1947, page 398.

9. The Electrical Conductivity of Living Tissues as it Pertains to Electrocardiography, **J. M. Benjamin, Jr., H. Schwan, C. F. Kay, J. H. Hafkenschiel.** *Circulation* (Philadelphia, Pa.), volume 2, 1950, page 321.

10. High-Fidelity Electrocardiography, **S. R. Gilford.** Conference on Electronic Instrumentation in Nucleonics and Medicine (New York, N. Y.), 1949.

11. The Value of High-Fidelity Electrocardiography Using the Cathode Ray Oscillograph and an Expanded Time Scale, **P. H. Langner, Jr.** *Circulation* (Philadelphia, Pa.), volume 5, 1952, page 249.

12. Standardization of Precordial Leads. *American Heart Journal* (St. Louis, Mo.), volume 15, 1938, page 107; volume 15, 1938, page 235, 1938; volume 25, 1943, page 535.

13. Electrocardiograms that Represent the Potential Variations of a Single Electrode, **F. N. Wilson, F. D. Johnston, A. G. Macleod, P. S. Barker.** *American Heart Journal* (St. Louis, Mo.), volume 9, 1934, page 447.

14. Experimental Studies on the Validity of the Central Terminal of Wilson as an Indifferent Reference Point, **M. Dolgin, S. Grau, L. N. Katz.** *American Heart Journal* (St. Louis, Mo.), volume 37, 1949, page 868.

15. La Vectocardiographie (book), **P. W. Duchosal, R. Sulzer.** S. Karger, New York, N. Y., 1949.

16. On Einthoven's Triangle, the Theory of Unipolar Electrocardiographic Leads and the Interpretation of the Precordial Electrocardiogram, **F. N. Wilson, F. D. Johnston, F. F. Rosenbaum, P. S. Barker.** *American Heart Journal* (St. Louis, Mo.), volume 32, 1946, page 277.

17. Fundamental Relations of the Instantaneous Electric Axis of Cardiac Accession, **R. H. Bayley.** *Medical Times* (East Stroudsburg, Pa.), volume 67, 1939, page 558.

18. On the Electrical Field Surrounding Doublets and Its Significance from the Standpoint of Einthoven's Equations, **R. Canfield.** *Heart* (New York, N. Y.), volume 14, 1927, page 102.

19. Heart Vector and Leads, **H. C. Burger. J. B. van Milaan.** *British Heart Journal* (London, England), part I, volume 8, 1946, page 157; part II, volume 9, 1947, page 154; part III, volume 10, 1948, page 229.

20. A Useful Network Theorem, **J. Millman.** *Proceedings,* Institute of Radio Engineers (New York, N. Y.), volume 28, 1940, page 413.

21. The Electric Field of an Eccentric Dipole in a Homogeneous Spherical Conducting Medium, **F. N. Wilson, R. H. Bayley.** *Circulation* (Philadelphia, Pa.), volume 1, 1950, page 84.

22. A Method of Analyzing the Electrocardiogram, **H. Mann.** *Archives of Internal Medicine* (Chicago, Ill.), volume 25, 1920, page 283.

23. The Determination and Significance of the Electric Axis of the Human Heart, **F. R. Dieuaide.** *Archives of Internal Medicine* (Chicago, Ill.), volume 27, 1921, page 558.

24. Aids for Determining Magnitude and Direction of Electric Axes of the Electrocardiogram, **C. E. Jackson, T. Winsor.** *Circulation* (Philadelphia, Pa.), volume 1, 1950, page 975.

25. A Slide Rule to Determine the Axis of the Electrocardiogram, **J. H. Beatty.** *Circulation* (Philadelphia, Pa.), volume 1, 1950, page 952.

26. A Study of the Relationship Between Unipolar Leads and Spatial Vectorcardiograms Using the Panoramic Vectorcardiograph, **W. R. Milnor, S. A. Talbot, E. V. Newman.** *Circulation* (Philadelphia, Pa.), volume 7, April 1953.

No Discussion

The RCA 3-Gun Shadow-Mask Color Kinescope

H. B. LAW
NONMEMBER AIEE

IT HAS long been apparent that the ideal reproducer for color television should consist of a single kinescope capable of displaying the color picture directly. Considerable interest in the problem has resulted in quite a number of proposals for single-tube color kinescopes, and much time and effort has been expended in experimental work. No systematic review of the literature will be given nor will a description of the operating principles of the various kinds of proposed tubes be undertaken, for an article covering these subjects has recently been published.[1] Rather, it is the purpose of this paper to give an over-all description of the Radio Corporation of America (RCA) 3-gun shadow-mask color tube including its construction and operating characteristics.

Before proceeding to the tube itself it may be well to review briefly a few fundamental concepts. To accomplish the objective of producing on the viewing screen of the receiver a full color image of the scene being televised, it is necessary not only to portray the relative brightnesses that occur in different areas of the scene but also the hue and saturation of these areas. Some of us are familiar with brightness from our black-and-white experience but are less familiar with the other two concepts, hue and saturation. The hue or dominant wave length of an area tells us whether its color is red, yellow, green, and so forth. The saturation of the color indicates the amount of white light that would have to be mixed with a pure spectrum color of the same hue as the area to match the color of the area. A spectrum color is 100 per cent saturated whereas white light has zero saturation. The saturation attribute of color is generally described by such terms as pale, pastel, deep, and so forth.

Brightness, hue, and saturation are psychological sensations. It is easily

---

Paper 53-175, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted April 25, 1952; made available for printing December 23, 1952.

H. B. Law is with the Radio Corporation of America, Princeton, N. J.

The author has drawn freely on published articles as well as unpublished reports. Furthermore, he gratefully acknowledges the full support and encouragement that has been given by Dr. D. W. Epstein and Dr. V. K. Zworykin.

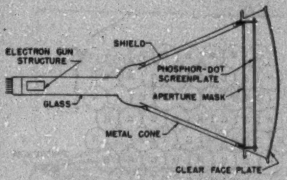


Figure 1. Cross-sectional diagram of the shadow-mask color kinescope showing major parts and their relative positions

possible to distinguish a large number of combinations of these sensations so that the reproduction of the brightness, hue, and saturation of each element of the scene to be transmitted would be an impracticable task if there were no relation between the stimuli. However, it has been found that color stimuli may be specified in terms of three known stimuli. These three are called primary stimuli, and since the picture is produced by colored light coming directly to the eye from the phosphors, the different colors are produced by the addition of the primary stimuli or, as sometimes stated, by an additive system of color. The primary stimuli are not unique, but are usually taken as red, green, and blue light in order to accommodate a large range of color. The procedure in specifying a color is to give the relative amount of each primary stimulus required to match the color. This is known as the tristimulus system of colorimetry or color specification. It makes possible the specification of the hue and saturation of each element of the scene to be televised by specifying the relative intensity of the red, green, and blue primaries needed to reproduce the color of the element. The brightness is reproduced by adjusting the sum of the intensities of the three primaries.

The shadow-mask tricolor kinescope employs additive color reproduction. Now, it may be said that the basic requirements of any additive color reproducer are:

1. A set of suitable primary lights.
2. A method for adding these primaries.
3. A method for controlling the amount of each primary to be added.

In regard to the first requirement, a set of suitable primary lights, the primaries that are used in the shadow-mask tube have their origin in the excitation of three kinds of phosphor that are placed on a glass plate inside the tube envelope. These phosphors have been chosen as having the best combination of high efficiency and desirable chromaticity that is available at the present time. The gamut of colors that it is possible to produce with these primaries compares very favorably with that possible with the best processes of color reproduction and is much superior to most commercial processes.

Turning now to the method of adding primaries, the system of juxtaposition is used in the tricolor tube. Small phosphor dots of three kinds are placed close together on a glass plate and three electron guns are located in such a position that all phosphor dots capable of emitting one of the primaries may be excited by electrons from one gun independent of the excitation of the remaining two kinds of dots. It is possible in the tricolor tube to use juxtaposition and obtain the effect of superposition. The eye blends the colors even at a close viewing distance because of the smallness of the phosphor dots.

The third requirement, a means of controlling the amount of primary light to be added at each point in the reproduced picture, may well start back at the camera. In the case of the RCA color camera, presently used, light from the scene to be televised is separated into red, green, and blue components and each component focused on the photocathode of a different image orthicon camera tube. The picture is then stored in the form of charge patterns on the target structures of the three image-orthicons until taken off by scanning beams traversing the targets in usual television fashion. The signal obtained from one of the camera tubes then contains the information to be fed to the particular gun in the tricolor tube controlling the excitation of the proper set of phosphor dots. When the output of the three camera tubes is properly fed to the grids of the three guns in the tricolor tube, the excitation of the three primaries in the tricolor tube will be of the proper amount at each point in the picture to reproduce the scene being televised.

Description of 3-Gun Tricolor Tube

Figure 1 is a cross-sectional diagram of the 16-inch color tube showing the major parts and their relative positions. The external metal envelope is made in two parts with flanges so that when placed together the two parts may be joined by a weld. The face plate of clear glass is sealed in place before the weld is made as is also the glass neck.

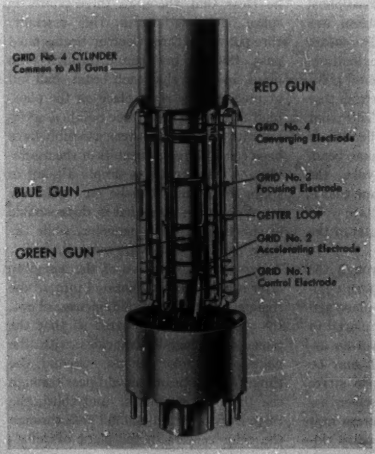


Figure 2 (left). Gun structure of the shadow-mask color kinescope

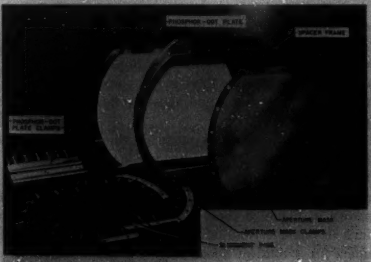


Figure 3 (right). Exploded view of viewing-screen assembly for the shadow - mask color kinescope

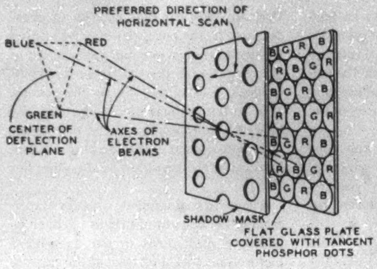

PREFERRED DIRECTION OF HORIZONTAL SCAN
BLUE
RED
GREEN
CENTER OF DEFLECTION PLANE
AXES OF ELECTRON BEAMS
SHADOW MASK
FLAT GLASS PLATE COVERED WITH TANGENT PHOSPHOR DOTS
B
G
R

Figure 4 (left). Diagram illustrating the principles of operation of the shadow-mask color kinescope

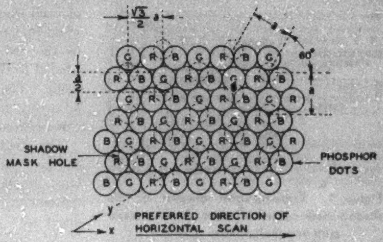


Figure 5 (right) Phosphor - dot array used in the shadow - mask color kinescope. The relation between the shadow-mask holes and the phosphor dots is shown for a region near the axis of the tube

Just inside the outer envelope there is fastened a high permeability metal cone to shield the electron beams from the earth's magnetic field as well as other stray magnetic fields.

Near the face plate is mounted the mask-phosphor screen assembly. It is for the purpose of inserting this assembly that the cone is constructed with flanges for a welded seal. The mask is on the gun side.

Mounted in the neck of the tube is the gun, Figure 2. There are three single parallel guns spaced equidistant from and at angles of 120 degrees about the assembly. The main lens for focusing the beam of each gun is between grids 3 and 4. The grid 4 cylinders of each of the guns open up into a single large cylinder, through which all beams travel. Grid 4 is called the converging electrode because, when the gun is in use, the lens action between the large cylinder and the conductive coating on the neck of the tube is used to converge the three beams on the aperture or shadow mask. Some focusing action on the three beams takes place, but the main function of the lens is to bring the beams together on the shadow mask.

Separate leads are brought out through a 14-pin base for the three cathodes and control grids. The three screen grids are also brought out separately to permit adjustment of individual drive characteristics. There is a common pin for the three first anodes, and the three heaters are connected in parallel to two base pins.

Figure 3 shows an exploded view of the viewing screen assembly for the 16-inch color tube. There are three major parts, the mask, spacer frame and phosphor dot plate.

The aperture mask is made of thin-sheet supernickel alloy, the area inside the border being covered with an array of regularly spaced holes. In the mask shown in Figure 3 these holes are 0.009 inch in diameter and spaced 0.023 inch apart for a total of about 195,000 holes in the picture area. There is an alignment hole and slot in the mask and a corresponding hole and slot in the spacer frame. The spacer frame provides the desired separation between mask and phosphor dot plate.

The glass plate which is to receive the phosphor screen is cut to match the outer contour of the frame and drilled near both ends of its horizontal center line to receive alignment pins that go through the alignment hole and slot in the spacer frame and mask.

Explanation of Tube Operation

As the name implies, the metal mask in front of the phosphor screen serves to mask or prevent the electron beams from reaching the phosphor screen except when the beams are going through the small holes in the mask. The successful operation of the masking principle for reproducing television pictures in color depends on being able to cause the individual electron scanning beams to strike one kind of phosphor only when they reach the phosphor screen. Solving the problem was made simpler because a scanning beam may, to a sufficiently good approximation, be considered as originating from a point and as traveling in straight lines from that point to the phosphor screen. This point may be called the center of deflection of the beam, and if more than one beam goes through the deflection yoke, the centers of deflection define a plane that is normal to the axis of the deflection yoke and lies somewhere near the center of the yoke.

If three beams in going through the yoke pass through the apexes of an equilateral triangle in the deflection plane and converge to a point on a mask placed in front of a phosphor screen, then an arrangement of holes in the mask may be found that permits each beam to strike separate areas on a phosphor screen beyond the mask such that the areas nest together perfectly. The geometry involved is shown in Figure 4, where the axis of the three beams passing through deflection centers are shown going through one hole of the mask and falling on phosphor dots in the plane of the phosphor screen. The holes in the mask are placed at the apexes of equilateral triangles that are oriented in such a way with respect to the deflection plane triangle that the phosphor dots, when in the appropriate places, all nest together perfectly. The relation between the mask holes and the phosphor dots is more clearly shown in Figure 5. The orientation of the beams in the deflection plane is not shown but may be visualized by imagining how the beams would have to go through the mask holes to hit the phosphor dots.

So far a geometrical arrangement has been indicated that makes possible the use of three beams and a shadow mask for separately exciting three types of phosphors. Assume for the moment that a shadow-mask phosphor-screen assembly has been built that fulfills the geometrical conditions just described. It will now be clear that there are three points in space associated with the assembly through which three electron beams must pass, that is, one beam through each point, in order that each beam shall fall at precisely the right place on the phosphor screen. These three points associated with the viewing-screen assembly have been called the color centers of the assembly. A color center is simply a point in space from which can be seen only one set of phosphor dots, that is, dots capable of emitting all the same color, while the other two sets of dots are masked.

For correct operation of the assembly after it has been incorporated into a tube there are then three requirements. First, the yoke should be located so that the plane of deflection coincides with the plane of the color centers. Second, the three electron beams should pass through the three color centers; and third, the three electron beams should pass through the color centers in the plane of deflec-

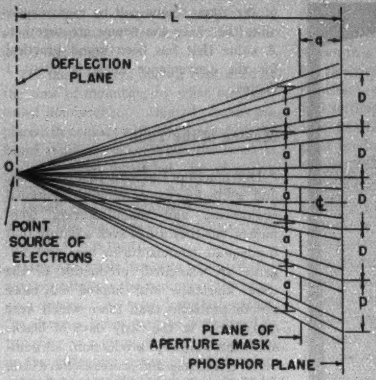


Figure 6 (left). A plane through an axis normal to the plane of the phosphor array and passing through a color center O in the shadow-mask color kinescope

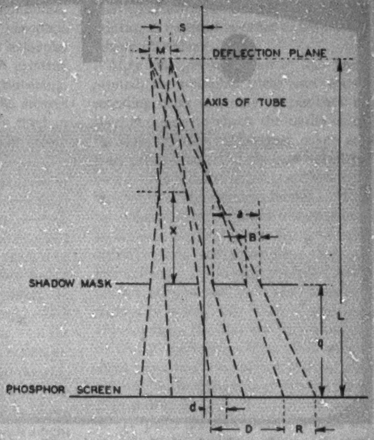


Figure 8 (right). A plane perpendicular to the shadow mask and passing through a row of holes in the y direction as defined in Figure 5. The geometry for a source of diameter M is shown

tion with the correct angles so that they are converged on the aperture mask.

First, the color-center plane and the deflection plane may be made to coincide by sliding the deflection yoke forward or back while looking at a single color field to determine the correct position for the yoke. If the second requirement is not fulfilled, which means that the beams do not go through the color centers because the gun mount is displaced from its correct position, then it is possible, by means of a small transverse magnetic field on the neck of the tube, to bend the beams until they do go through the color centers. The magnetic field is supplied by coils on either side of the neck of the tube and the coils may be rotated about the axis of the tube to obtain the proper direction of the field. This component is called a color-purifying coil.

The third requirement was that the beams should go through the color centers at the correct angle to converge at the mask. The three guns are mounted parallel to conserve space and facilitate accurate construction and alignment. The convergence lens bends each beam inwardly, but if slight misalignment of the guns makes it necessary, correct convergence may be obtained by adjusting the position of small magnets placed outside of the tube and adjacent to the guns so that nearly independent control over the three beams may be used to bring the beams together at the mask.

The geometrical properties of the viewing screen assembly are of sufficient importance to be considered a little further. The holes in the mask are placed at the apexes of equilateral triangles. There are, then, three directions in the pattern of holes in which well-defined rows of holes appear with the holes spaced apart a distance equal to the side of the equilateral triangle. Figure 6 shows a section through a color center and one of these rows of holes in the mask. The color center, O, is on the left and an edge view of the mask and phosphor screen is on the right. As indicated, the distance between the centers of the mask holes is a constant and here denoted by a. Also, the mask is parallel to the phosphor screen and is located at a distance q from the screen. If we consider the triangles made by drawing lines from the color center through the centers of all the mask holes to the phosphor screen, then it is evident that a number of triangles are formed all of which have the same altitude L. Consider now the upper triangle. By simple geometry, D over a is equal to L over $L-q$. The same proportion holds for each of the triangles and since a, L, and q remain constant, all the D's must be equal. The D's represent the center-to-center spacing of the correct phosphor dot positions, which means that the phosphor dot pattern consists of dots uniformly spaced from one side of the screen to the other. In fact, the phosphor dot pattern is the same as the mask pattern but enlarged by the factor L over $L-q$.

The practical significance of the fact that the phosphor positions are equally spaced for one of the colors is that if a stencil can be made for depositing the phosphor of one color, then the appropriate shift of the stencil will enable the phosphors of the second and third colors to be located because the three phosphor

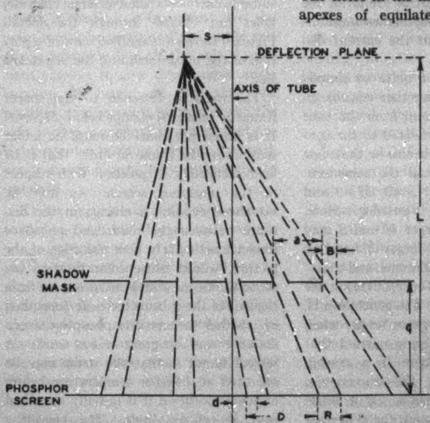


Figure 7. A plane perpendicular to the shadow mask and passing through a row of holes in the y direction as defined in Figure 5. The geometry for a point source is shown

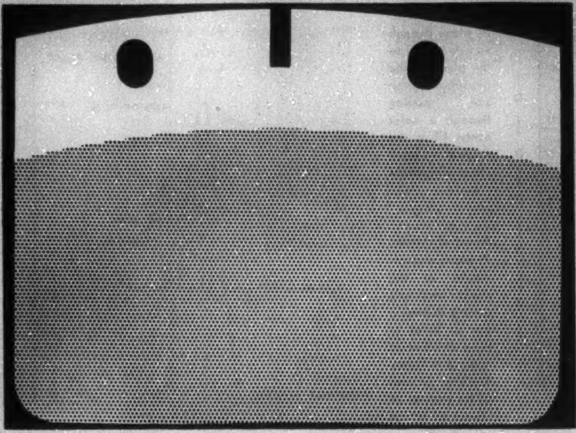


Figure 9. A magnified view of a portion of the aperture mask

patterns are identical and nest perfectly together.

So far, in discussing the geometry of the tube, each electron beam has been considered as originating from a point source at its center of deflection, as shown in Figure 7. This implies that the holes in the mask may be nearly equal in size to the phosphor dots. Since the electron beam at the center of deflection may have an appreciable diameter, the mask holes will have to be reduced from the size they could be in the point source case in order to obtain proper shadowing action at the phosphor screen. To study the effect, a simple approximation is to assume that electrons originate and travel in straight lines from a disk that represents the cross section of the beam at the center of deflection, Figure 8. Assume that peripheral rays define the electron beam as it goes through the shadow mask to the phosphor screen. In Figure 8 only one of the electron beams is shown together with one row of mask holes and the phosphor dots associated with that row. R is the phosphor dot diameter when the 3-color array consists of tangent dots, D is the distance between dots of the same color, a is the distance between aperture holes of diameter B. M is the diameter of the disk that is assumed to be the source of electrons, and the disk is located a distance S from the tube axis. Now the problem is to find the consequence of changing the beam width in the deflection plane from a point to some diameter M. Without going into geometrical proof it can be seen that if M goes to zero the mask hole diameter becomes a maximum. As M goes from a point to some value M, the peripheral rays require a smaller mask hole, and if the geometry were considered carefully it would be found that as the value of M got large enough so that this disk, together with disks for the other two beams not shown, became tangent, then the mask holes for proper shadowing would be zero. Zero mask hole diameter means zero electron current to the phosphor screen, while $M=0$ means zero current in the beam. There must, therefore, be an optimum value of M and B that will allow maximum beam current to get to the phosphor screen. The optimum values will, of course, depend on the current distribution across M, so an answer cannot be found for the size of M and B until further assumptions have been made about the current distribution in the beam.

The problem is not quite so simple, however, for there is another variable S, the distance of the beam from the tube axis. The value S is related to the convergence angle of the beams or the angle between each beam and the tube axis It can be seen that if $S=0$, $M=0$ and again no beam current is possible. However, as S becomes larger M and S may take on values. The larger S becomes the larger M or B may become, and consequently the more beam current may strike the phosphor screen. The conclusion is, then, that the convergence angle, when small, may limit the beam current that can be obtained. There is a reason, however, for keeping the convergence angle as small as possible. If a single deflection coil is to be used, the deflection of the three beams will be more nearly alike the closer the beams are together. A value that has been found practical for the convergence angle is about 1 degree.

Making of the Tube

Differences in the color tube and black and white kinescopes require additional engineering and manufacturing techniques, but they are of such a nature that if adequate manufacturing care and diligence are exercised, production of the color kinescope will impose no more serious problems than those which were experienced in the early days of black-and-white kinescope production. A number of techniques and processes have been worked out to produce shadow-mask color tubes.

The metal mask consists of a large number of holes in a sheet of thin metal with the holes uniform in size and equally spaced. Very satisfactory results have been obtained for pilot plant production of tubes by etching the holes in the masks from rolled metal strips. The alignment slot and hole as well as the border holes for mounting the mask are all made at the same time. In the process a photographic master is prepared having opaque dots where the holes are to be made in the mask. An ultraviolet light exposure through the photographic master or negative is then used to harden an acid resist that has been put on the metal. When developed, after the exposure, the resist washes off where there has been no exposure, or where the opaque dots were, while the space between the dots remains covered. By immersing the metal in suitable chemicals, such as ferric chloride, holes are etched through the metal. Figure 9 shows a magnified view of a portion of the mask including the alignment slot.

The mask is fastened to the spacer frame by means of clamps held by screws. It is important that the mask be under tension on the frame in order that it be held accurately in position with respect to the phosphor screen. As may be visualized readily, a change in the distance between the mask and phosphor screen, particularly near the edge of the picture, would cause color dilution because the beam would strike too far from the axis of the tube or too near depending on whether the mask to phosphor screen distance was too great or too small. A second factor is that the mask may be operated at a higher temperature before looseness occurs if it is initially stretched near its elastic limit. The operating temperature is somewhat higher than room temperature because of the beam power absorbed.

The mask tension may be controlled at the time of mounting by what has been called hot blocking. Heated platens are placed in contact with the mask, but the frame remains cool. While the mask is hot and thus expanded the clamping screws are tightened. When the mask is cooled it contracts and develops a tension that depends on the temperature to which it was raised.

After the mask has been mounted, the next operation is the making of the phosphor screen. As was described earlier, the electron beam may be considered as traveling in a straight line from a point in the deflection plane. It is therefore possible to simulate the electron beam by a point source of light and to locate the correct phosphor dot positions on a photographic plate placed in the plane where the phosphor screen is to be mounted. The apparatus used to perform these operations has been called the lighthouse. Figure 10 shows a schematic drawing that illustrated the apparatus. The mask-spacer frame assembly is placed with the mask down on locating points, not shown, for the purpose of accurately positioning the assembly with respect to the point light source below. A Kodalith plate is placed on the assembly with the emulsion towards the mask, so that the emulsion is in the same plane as the phosphor screen will be later. The point light source is located a distance below equal to the desired distance between the mask and deflection plane in the finished tube. Since the point source automatically fixes one of the color centers of the assembly at its own location when the exposure is made, it is important that the point source be positioned not on the axis of the assembly but off the axis a sufficient amount to give the required angle between the beam and the axis. The position of the point source about the axis is also important in order that the right angular relation between the mask array and the guns may ultimately be obtained.

As the exposure is made there are recorded on the Kodalith plate the positions where phosphor dots of a particular kind should be placed in order that all dots of that kind may be struck by the beam as it scans. At the same time, the alignment hole and slot are recorded on the plate as a guide to be used in positioning the phosphor plate when the phosphor dots have all been applied to form the screen.

From the Kodalith plate with the dots on it, or even directly from the lighthouse, one may proceed to make the phosphor screen in a number of ways, for example, by electrostatic printing, photo printing processes, offset printing, silk screening, and settling. However, only the process that has been most extensively used in pilot production, namely the silk screen process, will be described.

In general, the silk screen process of printing consists of blocking the openings to make a stencil out of a piece of cloth or metal mesh. The holes are blocked by a gelatin layer placed where there is to be no printing. Ink may then be forced through the clear parts of the fabric to the surface to be printed by means of a squeegee drawn across the fabric with the ink being pushed ahead of the squeegee.

The steps in preparation for printing the dot pattern are as follows: The Kodalith plate containing the dot pattern is lacquered and then covered with a light coating of paste wax such as Simoniz wax. A sheet of paper-backed photosensitized gelatin is applied to the emulsion side of the photographic plate. Ultraviolet light is then used to expose the gelatin with the dot pattern acting as a negative. During the development the unexposed areas dissolve out, leaving a gelatin stencil adhering to the photographic plate. The stencil consists of holes in the gelatin corresponding to the dots on the Kodalith plate. While still wet, the gelatin and plate are placed, gelatin side down, on a fine stainless steel wire screen stretched on a frame. The screen has wires 0.002 inch in diameter with 170 wires to the inch. After the gelatin is dry, the photographic plate is removed leaving the stencil adhering to the wire screen ready to use.

The actual printing is done on a table fitted with a vacuum mechanism for holding the glass to be printed. Also there is a crossfeed arrangement for moving the glass in two directions underneath the stencil, the amount of movement being indicated by dial gauges. The stencil is placed over the glass and its mechanism and held stationary. Movement of the glass with respect to the stencil is used for initial alignment of the pattern with the two holes in the glass as well as shifting the glass under the pattern to get the correct position for printing the second and third phosphor arrays. The movement for changing to another phosphor array is about 14 mils.

The printing ink or paste is made by dissolving ethyl cellulose in amyl alcohol and then adding the phosphor powder. Since the same stencil is used for printing each of the three types of phosphor, a thorough cleaning of the stencil after each printing is required in order to keep from mixing phosphors.

After one phosphor array has been printed, the phosphor plate and corresponding mask assembly are put back in the lighthouse to check the positions of the dots visually. The three colors are then printed and the plate is baked in air for 1 hour at 400 degrees centigrade to remove the ethyl cellulose binder. After the bake, a silicate solution is sprayed on to act as a binder to keep the dots from coming off in the subsequent steps of aluminizing. These steps consist of floating an organic film onto the phosphosurface in a tank of water, evaporating aluminum on the film and air baking to drive out the organic film. When these steps have been completed a thin, continuous film of aluminum covers the phosphor screen.

The parts are now ready for final assembly of the viewing screen. A steel alignment pin is pressed into the alignment hole in the spacer frame. The diameter of this pin is matched with that of the hole in the phosphor dot plate which is to fit over it. Next, an alignment bolt is fitted through the opposite hole in the phosphor dot plate and the slot in the spacer frame. Slippage along the horizontal axis, therefore, is permitted to the extent required by the differential expansion between glass and metal, which occurs during tube bake-out. It is to provide for this differential expansion that the slot is used, while insuring that the original alignment be restored after the heating cycles in making the tube are over.

The 3-gun assembly is sealed into the neck, the viewing-screen assembly is bolted to support posts in the lower cone and the upper cone put in place for the final welded seal. After the final seal the tube is exhausted by modified cathode-ray-tube methods. Reduced heating and

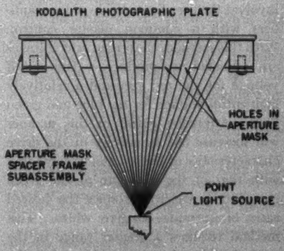

KODALITH PHOTOGRAPHIC PLATE
HOLES IN APERTURE MASK
APERTURE MASK SPACER FRAME SUBASSEMBLY
POINT LIGHT SOURCE

Figure 10. A schematic drawing illustrating how a light source is used to simulate one of the electron beams

cooling rates are desirable to avoid breakage of the phosphor dot plate from thermal shock and overstretching of the shadow mask due to too large a temperature differential between the mask and frame.

Operating Characteristics

In Figure 11 is shown the maximum operating voltage ratings. It will be noted that dynamic-focusing and convergence voltages are used. They consist of superimposed voltages that keep the beams focused and converged at all points on the flat phosphor screen.

For satisfactory performance, it is essential that there be no color shift with variation in intensity of light output. Since the phosphors show negligible current saturation in the range of current densities used, the necessary constancy of color requires that the ratios between the final electrode currents from the three guns be held constant over the desired brightness range. When the tube is used to reproduce black-and-white pictures, it is particularly important that the color temperature remain constant for the various shades of gray in the picture to prevent color tinting of the highlights and shadows.

To produce a white of approximately the same color temperature as that produced in a conventional black-and-white kinescope, it is necessary to have about three times as much current from the red gun as from the blue or green gun. One method of driving the three guns involves the use of separate video-gain controls on each of the amplifiers driving the control grids so that the drive can be adjusted as required. Higher drive for the red gun will be required with the present red phosphor. With this method, the control grid cutoff voltage is adjusted to about the same value for the three guns by varying the screen grid voltages.

Another method of driving the guns involves the use of equal-gain video amplifiers with a common master control. With this method, the screen grid voltages are adjusted to produce a low cutoff voltage for the red gun and a relatively high cutoff voltage for the blue and green guns. Adjusting cutoff in this manner will increase the sensitivity of the drive characteristics of the red gun as compared with that of the blue gun or the green gun, and, hence, will produce the desired ratio of currents to give white. This method requires a greater range in the control grid and screen grid supply voltages than the method utilizing separately adjustable driving voltages.

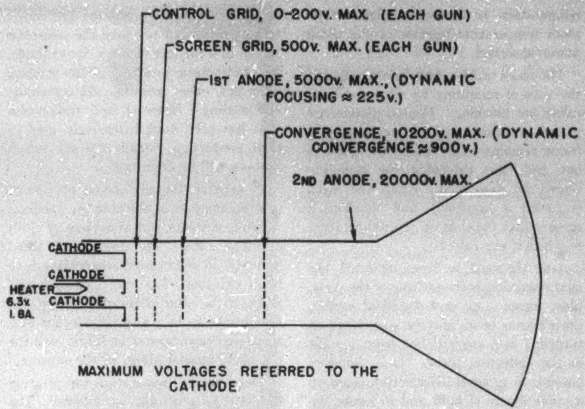

CONTROL GRID, 0-200v. MAX. (EACH GUN)
SCREEN GRID, 500v. MAX. (EACH GUN)
1ST ANODE, 5000v. MAX., (DYNAMIC FOCUSING ≈ 225v.)
CONVERGENCE, 10200v. MAX. (DYNAMIC CONVERGENCE ≈ 900v.)
2ND ANODE, 20000v. MAX.
CATHODE
CATHODE
HEATER
6.3v.
1.8A.
CATHODE
MAXIMUM VOLTAGES REFERRED TO THE CATHODE

Figure 11. Maximum operating voltages for the shadow-mask color kinescope

It was stated earlier that the mask operates somewhat above room temperature because of the beam power absorbed. For 20,000 volts on the final electrode an average current of the order of 300 microamperes may be used or of the order of 6 watts. This permits a peak current of 800 microamperes or more on average scenes. It has been found that under these conditions the temperature rise of the mask does not cause a shift of the mast by an amount sufficient to cause color dilution.

Highlight brightnesses of 50-foot lamberts have been obtained with the shadow-mask color tube, but on the average the tubes, are at this time operated at a somewhat lower brightness.

Turning now to the matter of resolution as influenced by the dot structure of the phosphor screen, there are perhaps two aspects to be considered. The first is the texture of the picture. To be sure, if one gets close enough to the phosphor screen the dot pattern can be seen. However, the structure is not visible at normal viewing distances. The other aspect is the resolution capabilities of the phosphor screen. Results obtained on a 16-inch tube indicate that substantially all the resolution can be seen on a test pattern that would be expected with a 4-megacycle bandwidth. Tests performed when no bandwidth limitation existed show that the tube is capable of a horizontal resolution of between 325 and 400 lines and a vertical resolution of between 350 and 450 lines. The fact that the vertical resolution is higher than the horizontal resolution is due primarily to the geometrical property of the pattern of dots. The resolution can of course be increased rather easily by increasing the number of dots. This has already been done in the 21-inch tube recently demonstrated in New York.

Conclusion

A detailed description of the RCA shadow-mask kinescope has been presented. In conclusion it may be emphasized that the technical developments leading to improved kinescopes, particularly since World War II, have formed the ground work to which has been added techniques particularly applicable to color tubes. A color tube has been built and put into pilot plant production that is capable of giving bright, well resolved pictures with excellent color separation. Each of the three primary images in the reproduced picture is separately produced by an electron beam. The use of a shadow mask insures that beam focus does not influence color while a single deflection yoke for the three beams eliminates the need for precisely linear scan.

Reference

1. METHODS SUITABLE FOR TELEVISION COLOR KINESCOPES, E. W. Herold. *Proceedings, Institute of Radio Engineers* (New York, N. Y.), volume 39, number 10, October 1951, pages 1177–85.

No Discussion

A New Design 60-Cycle A-C Network Analyzer

J. L. DAVIDSON
ASSOCIATE MEMBER AIEE

ROBERT E. KOLL
ASSOCIATE MEMBER AIEE

ANY power system, either present or proposed, may be reproduced in miniature and studied in careful detail with an a-c network analyzer. This device has become a powerful tool in designing and operating electric systems to cope with the rapid expansion of electrical facilities throughout the country. Its increased usage is reflected in the number of analyzers either built or planned, and in the waiting period of several months required to obtain a position on the schedule of those analyzers available for rental.

Conventional analyzers similar to those installed at the large electrical manufacturers' headquarters are expensive and, in general, can be justified by only the larger utilities and systems. A number of designs to achieve a lower cost device have been made or planned. Various approaches to low cost have included:

1. a 10,000-cycle design;
2. a 5-degree shift method to achieve low cost reactors;
3. a d-c method involving only resistors in conjunction with a special metering scheme;
4. the use of capacitors to replace reactors in a conjugate vector approach;
5. a successive-approximation solution of equations using an International Business Machines card punch calculating machine;
6. a number of homemade analyzers with fixed values of impedance to represent a fixed system.

Some of these methods are in satisfactory operation while others are in the development stage. References to published material on them are given at the end of this paper.

This paper describes a new design of low cost analyzer which has recently become available to the industry. It has been placed in successful operation by one of the smaller utilities having a peak load of approximately 500 megawatts and whose average rate of growth has exceeded 20 per cent per year over the past 5 years. The engineers of this company, as well as those of a number of other utilities, have worked very closely with the manufacturer to produce a flexible and conveniently operated design capable of wide application and usage.

This analyzer has a number of features which are not usually found in recent conventional designs. These features include:

1. operation of all elements at 60-cycle alternating current using moderate size reactors;
2. the use of variable autotransformers for generator voltage and phase-angle control;
3. load units which can be set directly as low as 1 per cent of board base on a single unit;
4. the use of continuously variable resistance and reactance units rather than tapped units with decade switches;
5. a variable reactance-over-resistance ratio adjustment for all generators;
6. load adjustment control which covers a complete voltage range;
7. calibration of resistance, reactance, and capacitance elements against a single decade resistance standard;
8. simultaneous metering of volts, amperes, watts, and vars;
9. accurate direct reading of I^2R and I^2X losses in any unit;
10. the use of solenoid-actuated precision snap switches for instrument selection circuits;
11. the use of a cathode-ray oscilloscope as a null detector in balanced bridge circuits for calibrating purposes.

The first three features mentioned are available in certain other recent nonconventional designs, while the remainder are believed to be unique in this design. In addition to these unusual features, this analyzer is equipped with many of the usual refinements found in most other modern designs such as low-burden electronic metering, push-button selection of any element to be metered, load-calibrating instruments, individual instruments with schedule-setting pointers for each generator, pilot-lamp indication of any element selected by push button, numerous transfer busses, and metered busses.

The total number of components placed in operation on the first board built to this design are:

12 generators (9 at 0 to 110 per cent X, 3 at 0 to 500 per cent X)
70 Pi lines (0 to 110 per cent R or X; susceptance 0 to 43 per cent each)
70 lines or loads (not available as lines when used as loads and vice versa)
35 tap changers (boost or buck 20 per cent; 1/2 per cent steps)
5 mutual transformers, 1-to-1 ratio
10 capacitor banks (0 to 103 per cent susceptance)
18 metered jumpers

These components are housed in a metering desk and nine cabinets and when assembled occupy a room 17 feet by 19 feet. The metering desk and typical cabinets are shown in Figures 1 and 2. Connections between components are made by external jumpers from the various sections to the bussing panels or to nearby sections.

Power Supply

A 3-phase 220-volt 60-cycle (or 50-cycle) a-c power supply is required. This voltage is applied to the machine through three delta-connected automatic voltage regulators which hold the voltage to within 1/2 volt. No auxiliary equipment is required for frequency multiplication. Total power requirements are approximately 600 watts.

Load and Line Units

Reactor size for a low-frequency analyzer is a problem because of the large size required for the high-impedance load units. By means of a variable autotransformer inserted ahead of a load unit, the reflected impedance of the load unit as seen by the autotransformer terminal will increase as the square of the transformation ratio. The connections of a load unit are shown in Figure 3. By proper choice of an impedance base, it is possible to use a moderate size reactor for either loads or lines. The resistor and reactor shown in Figure 3 may be independently used as a line unit when reconnected in series. The resistance elements consist of industrial-type wire-wound resistors, either fixed or variable. The reactor units are specially designed units with fixed taps and movable brushes, and hold a constant reactance over a wide range of current.

By metering at the point shown, the losses in the variable autotransformer and in the load-calibrating device are included as part of the total load. The var losses of the transformer are supplied by the small capacitor.

A typical panel containing two Pi lines

Paper 53-160, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted July 29, 1952; made available for printing December 16, 1952.

J. L. DAVIDSON is with the Long Island Lighting Company, Mineola, N. Y., and ROBERT E. KOLL is with the System Analyzer Corporation, Milwaukee, Wis.

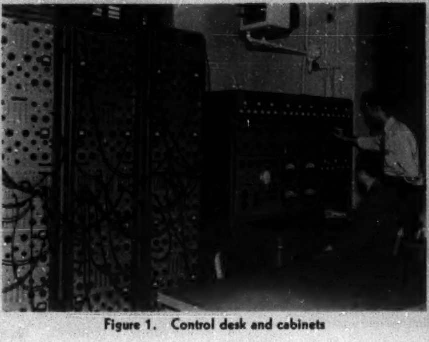


Figure 1. Control desk and cabinets

Figure 2. Cabinets located opposite control desk

and two loads is shown in Figure 4. The two Pi lines complete with capacitors are located in the top half of the panel. The two non-Pi lines are located in the bottom half of the panel with the variable autotransformers to convert them to load units immediately above. A tapped autotransformer is located at the bottom center of each of these panel sections. The reactors are arranged behind the panel so that there is no mutual coupling between units.

Generators

In the generator elements, use is again made of variable autotransformers.

These are used to achieve voltage regulation and phase-angle control. The connections, shown in Figure 5, give angle control over a 60-degree range. A complete rotation of 360 degrees is obtained by reconnecting to different phases by means of tap switches. Vernier control of both phase angle and voltage is achieved by a second brush on the variable autotransformer. This auxiliary brush supplies the primary winding of a small step-down transformer. The secondary winding is connected in series with the output terminals. Two concentric shafts are used to operate the two brushes independently. The vernier connections are shown only on the voltage control transformer to avoid confusing the diagram.

The generators are equipped with reactance units similar to those used with the line and load units. Every third generator has additional reactance to represent small machines. An adjustment is provided in each generator which allows the resistance-over-reactance ratio of the generator reactor to be set at any value from 0 to 5 per cent, or from a Q of infinity to 20.

A typical generator panel containing three generators is shown in Figure 6. Individual voltmeters, wattmeters, and varmeters with schedule-setting pointers are provided for each generator.

Calibrating Facilities

With all elements completely variable, it is necessary to adjust and calibrate against a known standard. This standard is a precision decade resistance box with a range from 0 to 10,000 ohms in steps of 0.1 ohm. The desired value is set on the decade resistance box and the unit selected is adjusted to give a null indication on a cathode-ray oscilloscope which is connected in an impedance bridge circuit. Separate bridge circuits are used for resistance, reactance, and capacitance.

Phase angles are read directly on a single-standard phase shifter which is adjusted for null detection on the oscilloscope with respect to the voltage under observation. The calibrating panel is shown in Figure 7 with the decade resistance standard at the top, the oscilloscope at the right, and the phase shifter at the lower left.

Metering Facilities

Four main instruments are provided for simultaneous indication of volts, amperes, watts, and vars. This panel is shown in Figure 8. These instruments are driven by highly stable electronic power amplifiers having a 35-decibel negative feedback. Input to the amplifiers is obtained from low-burden shunts and potential multipliers inserted in the desired metering location by precision snap switches. The instrument scales are hand-calibrated for each range of each instrument. Full-scale ranges are 1.5, 6, 25, 100, and 400 per cent base watts or vars. Full-scale voltmeter ranges are 9, 35, and 150 per cent. Full-scale ammeter ranges are 0.3, 1.2, and 4.8 per unit.

Push-button selection of any element to be metered is provided. The push-button switches actuate solenoid-driven cam shafts which in turn actuate the precision snap switches, each of which is permanently sealed in its own dust-tight housing. The cams are arranged to operate the switches as make-before-break contacts so that no current is interrupted by the metering bus contacts. Push buttons provide for metering of any terminal of any element, including the capacitors on the Pi sections. In addition, a simple switching device is provided to permit direct reading of I^2R and I^2X losses in any section.

The push-button panel with the metering range selector switches is shown in Figure 9. Selection of voltage and current ranges is done by tap switches. Signal lights shown in Figure 8 indicate the proper scales to be read on the watt and var instruments.

Bussing Facilities

The terminal of each generator appears at each cabinet. There are 12 local busses of 5 terminals each and 14 transfer busses of 3 terminals each in every cabinet. Three additional single-terminal transfer busses are also provided. All transfer busses appear in each cabinet. Tap switches for each load section permit that load to be connected to either terminal

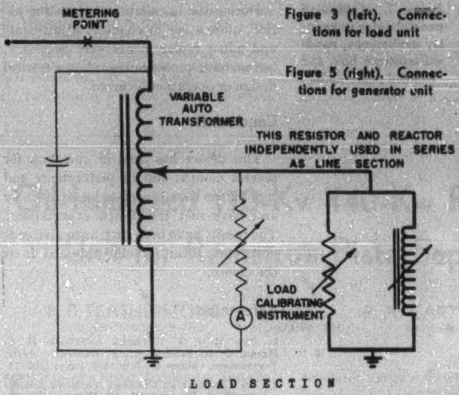
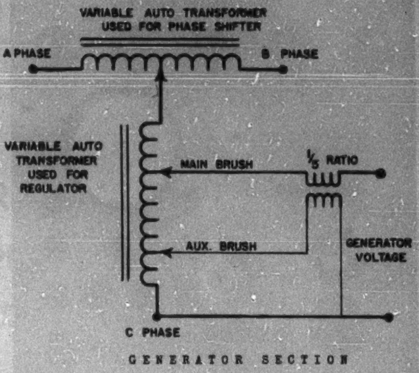

METERING POINT
VARIABLE AUTO TRANSFORMER
THIS RESISTOR AND REACTOR INDEPENDENTLY USED IN SERIES AS LINE SECTION
LOAD CALIBRATING INSTRUMENT
A
LOAD SECTION

Figure 3 (left). Connections for load unit

Figure 5 (right). Connections for generator unit


VARIABLE AUTO TRANSFORMER USED FOR PHASE SHIFTER
A PHASE
B PHASE
VARIABLE AUTO TRANSFORMER USED FOR REGULATOR
MAIN BRUSH
1/3 RATIO
AUX. BRUSH
GENERATOR VOLTAGE
C PHASE
GENERATOR SECTION

of the 2-line sections in the same panel. A view of a combined line and load panel is shown in Figure 4. The two Pi lines are at the top and the two load or line units are at the bottom. Three terminals are provided for each line section for the convenience of connections.

Tap Changers

A 20-per-cent boost-or-buck tap changer with a 1/2-per-cent steps is located in the center bottom of the panel shown in Figure 4. Here again, the var losses are supplied by a small capacitor. The watt losses are approximately 0.3 per cent of board base.

Capacitor-Mutual Panel

Ten capacitor banks with a total susceptance of 103 per cent each in 1-per-cent steps are available and may be used to represent either series or shunt capacitor banks on a system. Five 1-to-1 ratio mutual transformers are available on the same panel. Exciting-var losses of the mutual transformers are also supplied by small capacitors. This panel is shown in the upper right corner of Figure 2.

Base Quantities

The base quantities used on the analyzer are 50 volts, 1/2 ampere, 25 watts or vars, and 100 ohms. The meter scales are marked in per cent volts, per cent watts, per cent vars, and per unit amperes. The thermal capacity of all elements is equal to or greater than 500-per-cent base amperes.

Performance

A thorough check was made by comparing with typical load flow values obtained on other studies, and by checking against calculated values. In checking against results previously obtained on other analyzers, the division of loads among various lines was within a few per cent and voltages were quite accurate. The same applies to checks against known system conditions which were duplicated

Figure 4. Line and load panel showing two Pi lines and two load units. Loads may be used as normal lines. Tap changer is located in lower center

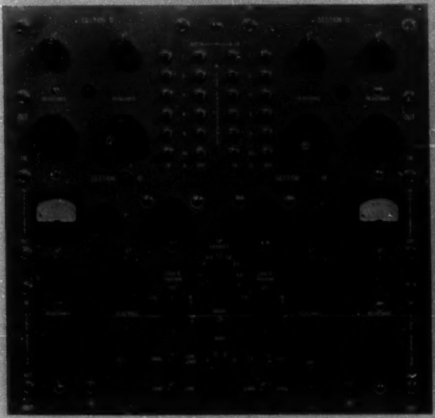
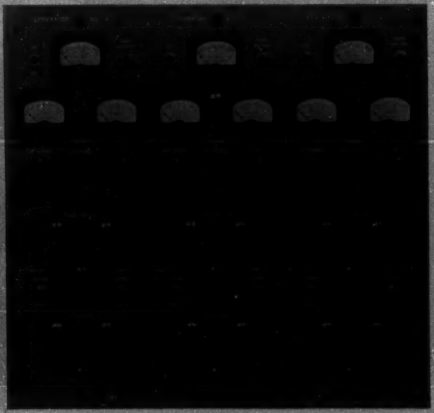


Figure 6. Generator panel showing three generator units



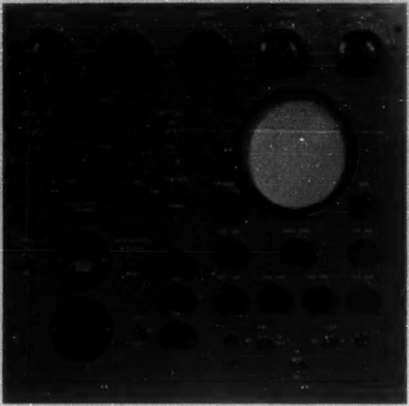


Figure 7. Calibrating panel showing cathode-ray oscilloscope, standard resistance box, and standard phase shifter

on the board.

A check against previously obtained transient-stability swing curves was also satisfactory.

A complete check of the board against itself was made regarding total system losses. Both watt and var losses were obtained by four methods:

1. calculating I^2R and I^2X from current readings and adding to obtain totals;
2. summation of I^2R and I^2X losses as read directly for each element;
3. summation of the in-minus-out differences of each element;
4. subtraction of the sum of the loads from the sum of the generation.

Assuming the I^2R and I^2X calculations from current readings to be 100 per cent, the direct reading method gave results within 1.4 per cent. The two latter methods ranged from 2.8 to 5.0-per-cent error. A larger error was to be expected for these methods since they both involve small differences of large quantities.

During the installation and testing period, quite a few mechanical adjustments and minor replacements were necessary, but no basic problems regarding electrical design or operation occurred.

Conclusion

This device has been in operation for several months and its performance and capabilities have proved to be satisfactory and have met the design expectations. The results have been such as to give confidence to future results obtained from the device.

References

1. THE M. I. T. NETWORK ANALYZER, **H. L. Hazen, O. R. Schurig, M. F. Gardner.** *AIEE Transactions*, volume 49, July 1930, pages 1102–14.

2. AN ALTERNATING CURRENT NETWORK CALCULATOR, **H. A. Travers, W. W. Parker.** *Electric Journal* (East Pittsburgh, Pa.), May 1930, pages 266–70.

3. A NEW A-C NETWORK ANALYZER, **H. P. Kuehni, R. G. Lorraine.** *AIEE Transactions* (*Electrical Engineering*), volume 57, February 1938, pages 67–73.

4. METHOD FOR A-C NETWORK ANALYSIS USING RESISTANCE NETWORKS, **Waldo E. Enns.** *AIEE Transactions* (*Electrical Engineering*), volume 61, December 1942, pages 875–80.

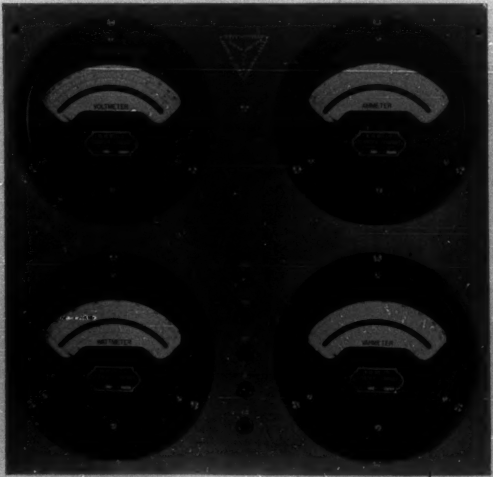
5. A NEW SIMPLE CALCULATOR OF LOAD FLOW IN A-C NETWORKS, **Waldo E. Enns.** *AIEE Transactions* (*Electrical Engineering*), volume 62, December 1943, pages 786–90.

6. A NEW DESIGN FOR THE A-C NETWORK ANALYZER, **J. D. Ryder, W. B. Boast.** *AIEE Transactions* (*Electrical Engineering*), volume 65, October 1946, pages 674–79.

7. THE USE OF BUSINESS MACHINES IN DETERMINING THE DISTRIBUTION OF LOAD AND REACTIVE

Figure 8 (below). Instrument panel

Figure 9 (right). Push-button panel with metering-range selector switches



Components in Power Line Networks, Philip D. Jennings, George E. Quinan. *AIEE Transactions*, volume 65, 1946, pages 1045–46.

8. Machine Computation of Power Network Performance, L. A. Dunstan. *AIEE Transactions*, volume 66, 1947, pages 610–20.

9. The General Solution Method of Power Network Analysis, Lyle A. Dunstan. *AIEE Transactions*, volume 67, part 1, 1948, pages 631–39.

10. A New Principle is Employed for 60-Cycle A-C Network Analyzers, E. B. Phillips. *AIEE Transactions*, volume 71, part 1, 1952, pages 18–25.

11. A Compact, Inexpensive A-C Network Analyzer, Edward W. Kimbark, James H. Starr, James E. Van Ness. *AIEE Transactions*, volume 71, part 1, 1952, pages 122–28.

No Discussion

Oil-Immersed 110-Kv 440-Kw Rectifier Unit for Resnatron Plate Supply

R. P. FEATHERSTONE
ASSOCIATE MEMBER AIEE

C. LING
NONMEMBER AIEE

P. A. CARTWRIGHT
MEMBER AIEE

THE proton linear accelerator installation at the University of Minnesota requires as one of its power supplies a d-c source capable of delivering continuously 4 amperes at 110 kv. This power represents the plate supply for four resnatron tubes which provide the radio-frequency power to excite the resonant cavities of the accelerator proper. The design of a power supply capable of meeting these requirements was complicated by the limited amount of available space.

Discussion of Components

To conserve floor space, a vault was built extending 9 feet below the floor level, a metal liner was provided, the vault partially filled with oil, and the rectifier components placed in it. This not only conserved space but was justified economically since all the transformers could be uncased and used without bushings. A 1-line diagram of the components involved in the power supply is shown in Figure 1. The primary service to the building's outdoor substation is 3-phase 4-wire 4,160 volts. Utilization voltage for the rectifier system is 480 volts and is supplied through a 600-ampere air circuit breaker, which is part of a metal-clad switchgear unit.

INDUCTION REGULATOR

The induction regulator used for voltage control of the rectifier is an oil-immersed, self-cooled, manually controlled, motor-operated unit. It is rated 250 kva, 480 volts; 100-per-cent raise and 100-per-cent lower. The regulator is provided with a neutral-stabilizing delta tertiary winding, and consists of three single-phase units mechanically connected together, mounted in a common tank, and driven by one operating motor. As is normally the case, the impedance of the regulator varies somewhat with the exact buck-boost position.

VAULT

The metal liner for the vault measures 17 feet by 7 feet by 5 feet in depth, is of 11-gauge cold-rolled steel, and has a capacity of 3,500 gallons of inhibited transformer oil. Provision is made for circulating the oil and for cooling it in a heat exchanger. The vault cover is made in two sections, one of which covers the rectifier transformers and the other the rectifier tubes and their filament supplies. Glass inspection ports are provided in the sections covering the rectifier tubes. As a safety measure a disconnect switch in the primary supply is provided in the cover, as can be seen in Figure 2.

The electric connections in the vault are made in such a manner as to allow individual removal of any one of the rectifier or filament transformers. The rectifier transformers are connected delta-delta with the deltas closed on the primary and secondary bus structures. Flexible leads and plug-type connectors allow the transformers to be disconnected from the busses and readily removed. Lowering and raising guides are provided to insure proper seating of each transformer unit, as shown in Figure 2.

TRANSFORMERS

The rectifier employs a conventional delta-delta 3-phase full-wave connection. Two factors favoring such a choice are the relatively high ratio of the d-c output voltage to the secondary rms voltage, and the possibility of operating open delta at reduced load in the event of a transformer failure. For very high voltage applications, a delta connection of the high-voltage winding is preferred to a wye connection because of its greater stability.[1,2] Each transformer is rated 167 kva with a primary voltage rating of 960 volts and a secondary voltage rating of 90,000 volts. Two primary voltage taps are provided, one at 5 per cent and one at 10 per cent below rated primary voltage, but at rated kva. The impedance of the transformers is approximately 16 per cent. The transformers were oil-impregnated by the supplier and shipped in a sealed, oil-filled tank. A portion of the transformers is shown in Figure 3.

The filament supply transformers for the rectifier tubes are rated 480 volts primary to 20 volts secondary. Each transformer has two secondary windings each rated at 20 volts, 32 amperes, to supply one of a pair of rectifier tubes. The transformer insulation is suitable for 150-kv operation. The tube sockets are mounted immediately above the transformers and are actually part of the transformer assembly. Figure 4 shows some details of this assembly.

RECTIFIER TUBES

The choice of a 3-phase full-wave bridge as the rectifier circuit required that there be six rectifier tubes, each having a peak inverse voltage rating of at least 130,000 volts, a peak current rating well above 4 amperes, and an average current rating of at least 1.3 amperes. No single kenotron was found which would meet these requirements, so series and parallel combinations of available tubes were considered. The final choice was the *ML-5576/200* kenotron, rated at 150,000 volts peak inverse, 2.5 amperes peak current, and 0.8 ampere average current, with two tubes in parallel at each rectifier position.

These kenotrons have an unusually low average tube drop of about 800 volts, less than 0.8 per cent of the power supply

---

Paper 53-67, recommended by the AIEE Electronic Power Converters Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted March 24, 1952; made available for printing December 2, 1952.

R. P. Featherstone, C. Ling, and P. A. Cartwright are with the University of Minnesota, Minneapolis, Minn.

This development was supported, in part, by the Atomic Energy Commission.

The authors wish to acknowledge the valuable assistance of Edward Day, Research Associate, University of Minnesota.

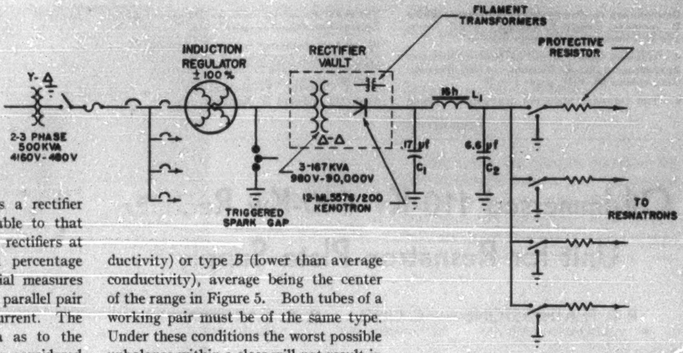


Figure 1. One-line diagram of 110-kv power supply

output voltage. This gives a rectifier efficiency which is comparable to that associated with mercury-arc rectifiers at lower voltages. The small percentage tube drop necessitates special measures to insure that each tube of a parallel pair will carry its share of the current. The manufacturer supplied data as to the range of tube characteristics considered acceptable by his quality control at final test; see Figure 5. It was evident that if a tube near the upper limit of conductivity were paralleled with one near the lower limit of conductivity, the former would be overloaded and would have short life.

A number of possible solutions to this problem were considered, such as adding sufficient resistance in series with each tube of a pair to insure acceptable current division or metering the current of each tube individually and adjusting variable resistors associated with each tube until equal currents were carried by each tube. The plan currently in use, however, is simply that of tube selection. Each tube is tested and classified as type *A* (higher than average conductivity) or type *B* (lower than average conductivity), average being the center of the range in Figure 5. Both tubes of a working pair must be of the same type. Under these conditions the worst possible unbalance within a class will not result in a serious overload for a tube. A simple test set, a bridge in which a tube is compared with a standard resistor, was built. It was found that tubes could be classified quite quickly and easily. Provision has been made for classifying into four, rather than two, subgroups if this should later prove desirable as a means of further improving the balance, and consequently, the tube life.

It is evident that if the filament of one tube of a pair were to open, the other tube would endeavor to conduct double current and would be dissipating many times rated wattage, because of the large voltage drop existing across a tube carrying an abnormal current. This requires that the system be shut down at once whenever any rectifier tube filament burns out in order to prevent damage to the good tube paired with it. A current-sensing relay is connected in series with the primary of each filament transformer, and is adjusted to drop out if the filament of either of the kenotrons supplied by that transformer burns out. Opening of any of the relays removes plate power from the system.

Because the life of rectifiers having pure tungsten filaments is critically dependent on filament voltage[3,4] it is advantageous to operate at reduced filament voltage when less than rated output current is required from the power supply, as will be the case a large part of the time in this application. Should the operator forget to increase the filament voltage when an increased output current is demanded there will be

Figure 2 (left). Rectifier vault prior to transformer installation

Figure 3 (right). Rectifier vault showing power transformers

Figure 4 (left). Rectifier vault showing tubes and filament transformers

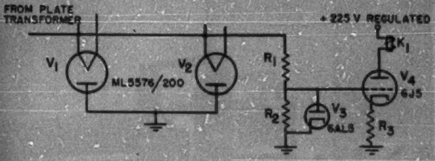


Figure 6 (above). Circuit for protection of ML-5576/200 kenotrons against excessive anode dissipation

an abnormally high tube drop and excessive anode dissipation. This will overheat the kenotron anodes and, depending on the severity of the overload, may damage them in a short time.

A simple means of avoiding damage of this type was adopted; it is shown schematically in Figure 6. Kenotrons V_1 and V_2 are one of the three pairs in the rectifier having grounded anodes. Care is taken to choose low-conductivity tubes for V_1 and V_2. The cathode-to-ground (and hence cathode-to-anode) voltage of the kenotrons is sampled by the resistance divider R_1-R_2, and the positive excursions due to inverse voltage are clipped by diode V_3. The negative excursions remaining are due to the forward voltage across the kenotrons. The constants of the rest of the circuit are so chosen that if the peak forward voltage exceeds a safe value the plate current of V_4 falls below the drop-out point of relay K_1. Opening of K_1 shuts off plate power to the system.

Filter Design

The filter consists of capacitors C_1 and C_2, and choke L_1, in the circuit diagram of Figure 1. Capacitor C_1 absorbs the energy of a commutating transient which, with only cable and stray capacities present, caused intolerably high ripple voltages (actually repeated damped oscillations superimposed on the normal ripple voltages) at the input to the filter. Because of the relatively high reactance of the plate transformers, the presence of this small input capacitor does not increase the peak current drawn through the rectifier tubes significantly above the choke-input condition. The size of C_1 is not critical and was determined experimentally.

Capacitor C_2 is an energy reservoir for pulsed operation of the resnatrons. The linear accelerator requires such large amounts of radio-frequency power that it must be operated in pulses, at a relatively low duty, to keep the energy requirements and the cooling problem within reasonable bounds. The pulse length is 300 microseconds, and the pulse repetition rate is 60 per second. From conservative estimates of the probable radio-frequency power requirements and of the probable plate efficiency of the resnatrons, it was calculated that capacitor C_2 must be able to supply an average power during the pulse of 24,000 kw.

The larger C_2 is chosen, the more nearly constant the voltage across it will remain during the pulse, but at the same time the stored energy will be increased so that greater damage may be inflicted on a resnatron in the event of an accidental arc-over within the tube. The resnatron manufacturer specified that C_2 be selected such that the voltage across it would not drop more than 10,000 volts during the pulse. From these considerations the required capacitance of C_2 is determined to be 6.52 microfarads.

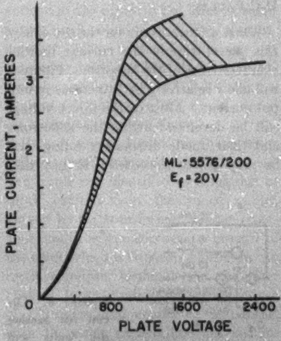


Figure 5. Acceptable range of characteristics of ML-5576/200 kenotron

Since the capacitors available to us were rated at 1 microfarad and 25 kv, the bank C_2 is designed by using five such units connected in series to stand the total voltage and 33 in parallel to make a total capacitance of $33/5 = 6.6$ microfarads. Figure 7 shows the construction of C_2. The division into groups with paralleling only at the top and bottom is done to limit the energy which would be dissipated within any one capacitor in the event it failed.

The filter choke L_1 smooths the current flow which the rectifier supplies to energy reservoir C_2. If there were no choke in the circuit, the rectifier would supply very large currents to C_2 immediately after the pulse, requiring rectifier tubes of uneconomically large peak current capacity. The calculations by which the required value of L_1 was determined are given in the Appendix. The minimum value was calculated to be 10 henrys.

Protective Resistors

This rectifier is to supply plate power for a group of resnatrons which, like other continuously pumped vacuum tubes of high power, are expected to spark over internally from time to time. Because the main filter capacitor will normally have about 40,000 joules of stored energy which will be discharged in any such spark-over, considerable damage to the resnatron tube structure is possible and special measures to reduce the energy delivered to the tubes under fault conditions are required.

The protective technique used is shown in Figure 1. A resistor is inserted in series with the output to each resnatron. It is chosen to be small by comparison to the effective tube anode impedance in normal operation (about 5 per cent) but large by comparison to the impedance of an arc. The resistors are noninductively wound (Aryton-Perry winding), with bare Nichrome wire on grooved insulating forms. They dissipate approximately 5 per cent of the total output power of the

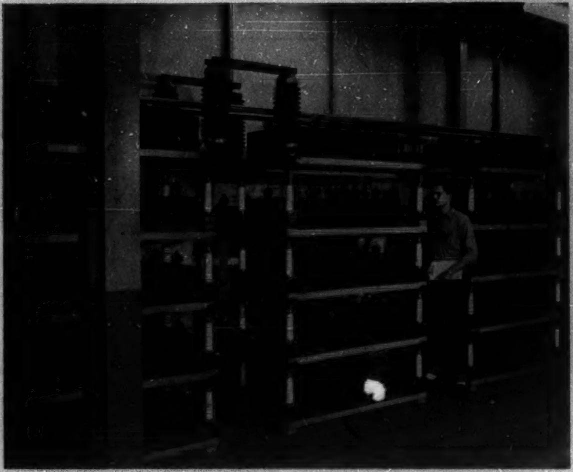


Figure 7. Construction of 6.6-microfarad 125-kv capacitor bank

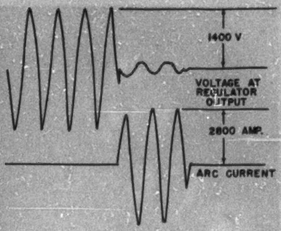


Figure 9. Oscillogram of typical spark gap operation

power supply in normal operation, and are able to withstand full supply voltage from end to end briefly under fault conditions. Any single resistor is able to absorb the stored energy of the capacitor bank without undue temperature rise.

Distribution

The high-voltage cable used for distribution of the vault output is of the type used for X-ray applications and is rated for service at 140 kv d-c. It has an overall diameter of 1.745 inches and a minimum bending radius of five times the diameter. The cable was supplied in six ready-cut terminated sections. The bushing used for the feed from the vault is of the 46-kv class; Figure 4.

System Protection

Fire Protection

Fire protection for the rectifier vault is provided by means of a carbon dioxide system which will blanket the vault area both above and below the vault covers. Temperature sensing units for automatic operation are provided both above and below the covers and the system can be triggered manually from the linear accelerator control room.

Circuit Breakers

As indicated in Figure 1, the rectifier installation is supplied from a 480-volt 600-ampere air circuit breaker with an interrupting rating of 25,000 amperes. This breaker is operated in cascade with the main breaker of the metal-clad switchgear unit. The main breaker has a continuous current rating of 1,200 amperes, and 50,000 amperes interrupting capacity. All the breakers of the metal-clad unit are electrically operated.

D-C Faults

When a fault occurs on the d-c side of this power supply the current limiting characteristics of the kenotrons, Figure 5, will allow relatively little increase in output current. Most of the direct voltage will be developed across the kenotrons, and their anode dissipation ratings will be very greatly exceeded. It may take as much as 5 cycles (1/12 second) for the d-c overload relay to operate and the main circuit breaker to clear the fault, and it is possible for the kenotrons to be seriously damaged in this interval.

A faster means of interrupting or reducing a-c power input was therefore sought. It seemed possible that short-circuiting the supply line from the induction regulator to the rectifier transformers could be done rapidly, would reduce a-c power input to a level which would be safe for the kenotrons during the circuit-breaker clearing time, and would not be excessively hard on the induction regulator and circuit breaker because of current limiting by the reactance of the induction regulator.

It was found that a 3-phase arc could be struck reliably and almost instantly by the circuit, shown in Figure 8. Abnormal current flow in the ground return of the main filter capacitor C_2 causes the thyratron to fire, discharging capacitor C_3 into the primary of the pulse transformer. The high-voltage impulse at the secondary of the pulse transformer is coupled to the initiating electrode of the spark gap, and sparks jump from this electrode to the three line electrodes. For reliable operation, it is essential that all three line electrodes be coupled by arc streams, because the striking impulse lasts for only a few microseconds; if only two line electrodes are short-circuited together, it may occur at a time when the voltage difference between them is too low to start and maintain an arc. The three small inductors shown in series with the phase leads present a high impedance to the steeply rising striking impulse so that the voltage continues to rise even after some of the gaps have broken down. In this way a single striking pulse breaks down all the gaps and all three line electrodes are coupled by arc streams. Once the 3-phase arc strikes, it will sustain itself reliably at normal operating conditions.

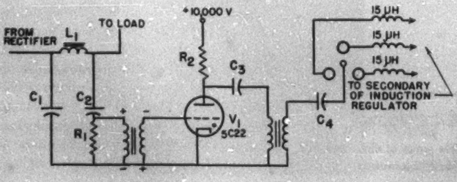


Figure 8. Circuit for sensing d-c faults and triggering 3-phase spark gap

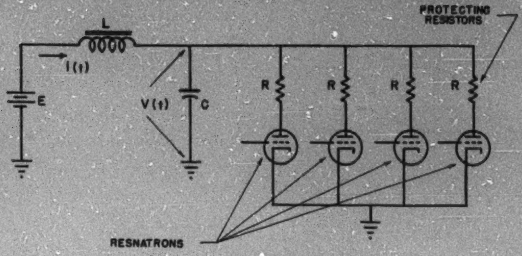


Figure 10. Equivalent rectifier circuit

The main electrodes used are 3/4-inch diameter carbon rods sheathed with copper to minimize resistive drop. The gap spacing used it 1/32 inch at initial adjustment; the gaps are reset after eroding to about 1/16 inch, which occurs after eight operations at full power. Longer gaps could be used except for the fact that the power supply may be used for considerable periods at reduced voltage, in which case the voltage available to maintain the arc is correspondingly reduced. The shorter the gap, the lower the voltage at which the spark gap is able to operate reliably.

Figure 9 is an oscillogram of a typical spark gap operation. The upper trace is the voltage at the rectifier-transformer primary terminals; the lower trace is the current drawn by one arc electrode. It can be seen that the voltage drops immediately from full value to about 100 volts rms, remaining at about the same level for 3 cycles, when the main circuit breaker clears the fault. It has been calculated that the temperature of the kenotron anodes will increase less than 100 degrees centigrade during such a fault, an easily tolerable increase.

Testing and Test Results

TESTS

When the construction of the rectifier was completed a series of tests was performed to determine its operating characteristics. The testing proceeded in these steps:

1. No-load test with no output filter.
2. No-load test with output choke only.
3. No-load test with complete output filter assembly.
4. Load test without filter.
5. Load test with choke only.
6. Load test with choke and capacitor assembly on the output.
7. Open-delta tests.

For purposes of testing under load the rectifier was connected to a water resistor consisting of tap water flowing in four parallel porcelain tubes.

For all tests, the following values were measured and recorded:

1. Output current and voltage.
2. Input current and voltage.
3. Current in the delta primary.
4. Oscillograms of the quantities tested above. For the oscillograms, a Polaroid land camera was used and proved very expedient for this type of work.

TEST RESULTS

The system was tested at no-load voltages as high as 137,000 volts and at currents as high as 4.3 amperes with satisfactory operation. It was run continuously for 7 hours at an output of 4.25 amperes and 120,000 volts, to check heating of components. It will be put into service when the resnatron installation is completed.

During initial tests, an unexpected difficulty was encountered with high ripple voltages at the input to the filter. The filter was initially designed as a conventional choke-input single-L section. The calculated peak-to-peak ripple voltage at the input to the filter at rated output with resistive load, was approximately 17,000 volts. Under test at reduced output, ripple voltages were observed equivalent to 80,000 volts peak-to-peak at full output, which would exceed the rating of the filter choke.

By experimental work on the rectifier and on a small scale model of it, the excessive ripple was identified as being due to shock excitation of the tuned circuit formed by the transformer leakage reactance and the capacitance to ground of the cable which is connected between the rectifier output terminal and the filter choke. It was found that a capacitance of 1/6 microfarad connected in parallel with the cable capacitance did not increase the peak currents which the rectifier tubes must handle, and reduced the ripple voltage at the input to the filter choke to less than 5,000 volts peak. This capacitor was incorporated in the filter design as C_1, Figure 1.

Conclusions

It is the opinion of the authors that a rectifier such as described in this paper is somewhat unique in its field. Many high-voltage supplies have been built but very few with a relatively high current rating. The space-saving features embodied in this design would be useful in any installation where space is at a premium.

Preliminary testing of this rectifier has indicated that the design is basically sound, and there have been no troubles to date with the standard equipment used. Little difficulty has been experienced with the items developed specifically for this application. Future testing may result in minor alterations of the components, but no basic changes in the overall design are contemplated.

Appendix. Calculation of Required Value of Choke Inductance L_1

To simplify the mathematical analysis, the following assumptions are made:

1. The output of the rectifier is a constant direct voltage source, that is, free of ripples.
2. The impedance of the source is zero.
3. The resistance of the choke is zero.
4. The pulse duration (300 microseconds) is negligibly short compared to the pulse interval. See Figure 11.

The energy fed by the capacitor bank to the resnatrons is stored during the charging period from the power supplied through the rectifier tubes and the filter choke L_1. The average charging current flow through the rectifier tubes and the choke L_1 is dependent only on the power consumed by the resnatrons during the pulse

$$i_{av}=\frac{24\times10^6\times0.018}{110\times10^3}=3.92\text{ amperes}$$

The maximum (or peak) current flowing through the rectifier tubes will be varied for different values of choke L_1; for two *ML-5576/200* rectifier tubes in parallel, a maximum current of 5 amperes is allowable.

In Figure 10, using Kirchoff's Law

$$E=L\frac{di}{dt}+\frac{1}{C}\int^t idt+Vc_0 \tag{1}$$

Differentiating with respect to time

$$L\frac{d^2i}{dt^2}+\frac{i}{C}=0 \tag{2}$$

Solving

$$i(t)=A\epsilon^{j\omega t}+B\epsilon^{-j\omega t} \tag{3}$$

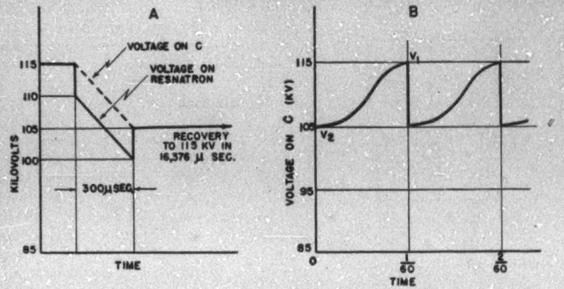


Figure 11(A). Output voltage wave forms—expanded time axis. (B). Output voltage wave forms

where

$$\omega = 1/\sqrt{LC}$$

Let $t=0$ be the time at which the capacitor is discharged to V_2 in Figure 11(B). At this instant the charging current through L_1 is $i=i_0$. Put in equation 3, then

$$i_0 = A+B \tag{4}$$

From equation 1 at $t=0^+$

$$\frac{1}{C}\int_0^t i\,dt = 0$$

so

$$L\frac{di}{dt} = L(j\omega A\epsilon^{j\omega t} - j\omega B\epsilon^{-j\omega t}) = E - V_2$$

At $t=0$

$$A - B = \frac{E-V_2}{j\omega L} \tag{5}$$

Solving for A and B

$$A = \frac{1}{2}\left(i_0 + \frac{E-V_2}{j\omega L}\right)$$

$$B = \frac{1}{2}\left(i_0 - \frac{E-V_2}{j\omega L}\right)$$

$$i(t) = \frac{1}{2}\left(i_0 + \frac{E-V_2}{j\omega L}\right)\epsilon^{j\omega t} + \frac{1}{2}\left(i_0 - \frac{E-V_2}{j\omega L}\right)\epsilon^{-j\omega t}$$

$$= i_0\left(\frac{\epsilon^{j\omega t}+\epsilon^{-j\omega t}}{2}\right) + \left(\frac{E-V_2}{\omega L}\right)\times\left(\frac{\epsilon^{j\omega t}-\epsilon^{-j\omega t}}{2j}\right)$$

$$= i_0\cos\omega t + \left(\frac{E-V_2}{\omega L}\right)\sin\omega t$$

$$= \sqrt{\left(\frac{E-V_2}{\omega L}\right)^2 + i_0^2}\,\sin(\omega t+\phi) \tag{6}$$

where

$$\phi = \tan^{-1}\frac{i_0\omega L}{(E-V_2)}$$

Let σ be the charging period. Then there will be two cases:

Case 1: $\sigma \leq \pi/\omega$

Figure 12 shows the variation of current through the choke and the voltage across the capacitor with time. The average, maximum, and rms currents will next be calculated.

$$\sigma + \frac{2\phi}{\omega} = \frac{\pi}{\omega};\quad \phi = \frac{\pi}{2} - \frac{\omega\sigma}{2}$$

$$\tan\left(\frac{\pi}{2} - \frac{\omega\sigma}{2}\right) = \frac{i_0\omega L}{(E-V_2)},\ \text{or}\ \cot\frac{\omega\sigma}{2} = \frac{i_0\omega L}{(E-V_2)}$$

So

$$i_0 = \frac{(E-V_2)}{\omega L}\cot\frac{\omega\sigma}{2} \tag{7}$$

Putting the value of i_0 in equation 6

$$i(t) = \sqrt{\left(\frac{E-V_2}{\omega L}\right)^2 + i_0^2}\,\sin(\omega t+\phi)$$

$$= \frac{(E-V_2)}{\omega L}\sqrt{1+\cot^2\frac{\omega\sigma}{2}}\,\sin(\omega t+\phi)$$

$$= (E-V_2)\omega C\csc\frac{\omega\sigma}{2}\sin(\omega t+\phi) \tag{8}$$

Maximum current occurs at $t=\sigma/2$.

$$i_{\max} = (E-V_2)\omega C\csc\frac{\omega\sigma}{2}\sin\left(\frac{\omega\sigma}{2}+\frac{\pi}{2}-\frac{\omega\sigma}{2}\right)$$

$$= (E-V_2)\omega C\csc\frac{\omega\sigma}{2} \tag{9}$$

To obtain the average current

$$i_{av} = \frac{1}{\sigma}\int^{\sigma} i\,dt$$

$$= \frac{(E-V_2)\omega C\csc\frac{\omega\sigma}{2}}{\sigma}\int_0^{\sigma}\sin(\omega t+\phi)\,dt$$

$$= \frac{-(E-V_2)C\csc\frac{\omega\sigma}{2}}{\sigma}\cos(\omega t+\phi)\Big|_0^{\sigma}$$

$$= \frac{2(E-V_2)C}{\sigma} \tag{10}$$

The rms current is

$$i_{rms} = \sqrt{\frac{1}{\sigma}\int_0^{\sigma} i^2\,dt}$$

$$= \sqrt{\frac{1}{\sigma}\int_0^{\sigma}\left[(E-V_2)\omega C\csc\frac{\omega\sigma}{2}\right]^2\times\sin^2(\omega t+\phi)\,dt}$$

$$= \frac{(E-V_2)\omega C\csc\frac{\omega\sigma}{2}}{\sqrt{2\sigma}}\sqrt{\left(\sigma+\frac{1}{\omega}\sin\omega\sigma\right)} \tag{11}$$

The following relations may be written down:

$$\frac{i_{\max}}{i_0} = \frac{(E-V_2)\omega C\csc\frac{\omega\sigma}{2}}{(E-V_2)\omega C\cot\frac{\omega\sigma}{2}} = \sec\frac{\omega\sigma}{2} \tag{12}$$

$$\frac{i_{\max}}{i_{av}} = \frac{(E-V_2)\omega C\csc\frac{\omega\sigma}{2}}{\frac{2(E-V_2)C}{\sigma}} = \frac{\omega\sigma}{2}\csc\frac{\omega\sigma}{2} \tag{13}$$

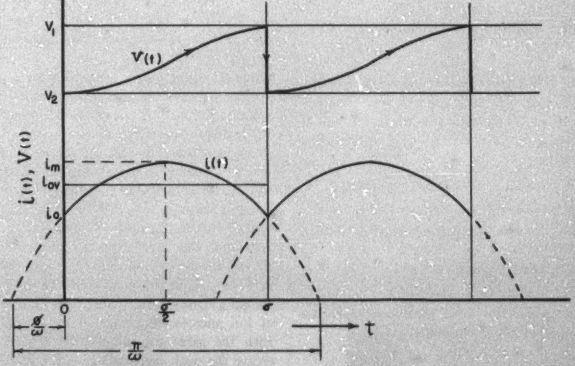


Figure 12. Choke current, and capacitor voltage for case 1

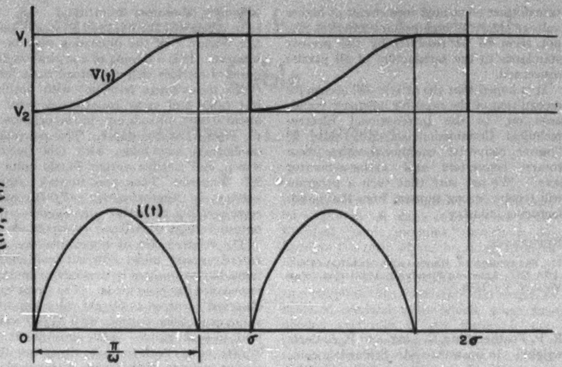


Figure 13. Choke current, and capacitor voltage for case 2

The voltage on C_2 at time $t=\sigma$ is

$$V_1=\frac{1}{C}\int_0^{\sigma} idt+V_2$$

$$=(E-V_2)\omega \csc\frac{\omega\sigma}{2}\int^{\sigma}\sin(\omega t+\phi)dt+V_2$$

$$=(E-V_2)\csc\frac{\omega\sigma}{2}\;[\cos(\omega t+\phi)]_0^{\sigma}+V_2$$

$$=2(E-V_2)+V_2=2E-V_2$$

So

$$E=\frac{V_1+V_2}{2} \qquad (14)$$

CASE 2: $\sigma>\pi/\omega$.

For this case, the graph of charging current and capacitor voltage versus time would appear as in Figure 13.

$$i_0=0$$

$$i(t)=A\sin\omega t,\; t\leq\frac{\pi}{\omega} \qquad (15)$$

$$i(t)=0,\quad \frac{\pi}{\omega}<t<\sigma$$

The constant A is again found from the relation

$$L\frac{di}{dt}+\frac{1}{C}\int_0^t idt=E-V_2$$

At $t=0$

$$E-V_2=L\frac{di}{dt}=LA\omega\cos\omega t=LA\omega$$

So

$$A=\frac{E-V_2}{\omega L}$$

$$i(t)=\frac{(E-V_2)}{\omega L}\sin\omega t \qquad (16)$$

for $t\leq\frac{\pi}{\omega}$

Maximum current is simply

$$i_{max}=\frac{(E-V_2)}{\omega L}=(E-V_2)\omega C \qquad (17)$$

The average current is

$$i_{av}=\frac{1}{\sigma}\int_0^{\pi/\omega} idt$$

$$=\frac{1}{\sigma}\int_0^{\pi/\omega}(E-V_2)\omega C\sin\omega t dt$$

$$=\frac{2(E-V_2)C}{\sigma} \qquad (18)$$

The rms current is

$$i_{rms}=\sqrt{\frac{1}{\sigma}\int_0^{\pi/\omega}(E-V_2)^2\omega^2C^2\sin^2\omega t dt}$$

$$=(E-V_2)C\sqrt{\frac{\omega\pi}{2\sigma}} \qquad (19)$$

Then

$$\frac{i_{max}}{i_{av}}=\frac{(E-V_2)\omega C}{\frac{2(E-V_2)C}{\sigma}}=\frac{\omega\sigma}{2} \qquad (20)$$

As before, the voltage across the capacitor just before discharge is

$$V_1=\frac{1}{C}\int_0^{\pi/\omega}(E-V_2)\omega C\sin\omega t dt+V_2$$

$$=2(E-V_2)+V_2$$

$$=2E-V_2$$

So

$$E=\frac{V_1+V_2}{2} \qquad (21)$$

In the design, $i_{av}=3.92$ amperes; $i_{max}=4.7$ amperes (for the reason of leaving some safety margin on the ripples neglected in the analysis, it is chosen to use $i_{max}=4.7$ amperes instead of 5 amperes). First, check with case 2: $\sigma>\pi/\omega$

$$\frac{i_{max}}{i_{av}}=\frac{4.7}{3.92}=1.2,\; \sigma=\frac{1}{60}=0.0167$$

Using equation 20

$$\omega=\frac{i_{max}2}{i_{av}\sigma}=1.2\times2\times60=144$$

Then

$$\frac{\pi}{\omega}=\frac{\pi}{144}=0.0218$$

So

$$\sigma<\frac{\pi}{\omega}\text{ (contradictory)}$$

Therefore we have to use case 1: $\sigma\leq\pi/\omega$ Using equation 13

$$\frac{i_{max}}{i_{av}}=1.2=\frac{\omega\sigma}{2}\csc\frac{\omega\sigma}{2} \qquad (22)$$

By cut and try method, we find $\omega=123.8$. Then

$$L=\frac{1}{\omega^2C}=\frac{1}{(123.8)^2\times6.6\times10^{-6}}=10\text{ henrys}$$

The output voltage of the rectifier circuit should be maintained at

$$E=\frac{V_1+V_2}{2}=\frac{115+105}{2}=110\text{ kv}$$

References

1. HIGH-VOLTAGE RECTIFIER CIRCUITS: PART II, E. V. DeBlieux. *General Electric Review* (Schenectady, N. Y.), volume 51, number 3, March 1948, pages 31–35.

2. HIGH-VOLTAGE STEEL-TANK MERCURY-ARC RECTIFIER EQUIPMENT FOR RADIO TRANSMITTERS, PART II, J. C. Reed. *Journal*, Institution of Electrical Engineers (London, England), October 1945, page 453.

3. RULES FOR PROLONGING TUBE LIFE, Hampton J. Dailey. *Electronics* (New York, N. Y.), volume 16, April 1943, pages 76–78.

4. REFERENCE DATA FOR RADIO ENGINEERS (book). Federal Telephone and Radio Corporation, New York, N. Y., second edition, 1946, pages 132–33.

◆

Discussion

R. G. Schlawin (General Electric Company, Pittsfield, Mass.): The authors are to be commended for shedding a lot of light on a quite dark subject. Their treatment of the general subject of a d-c power supply and the new application to resnatron tubes is very instructive.

The short-circuiting system for reducing duty on rectifier tubes has not been otherwise covered in the literature to my knowledge. This method, colloquially known to the people working in nucleonics as the "Crowbar," is very important to the proper operation of the system. It is fortunate that this knowledge is now becoming available to others besides nucleonics specialists.

For many applications, an extra power loss of 5 per cent of the power output during normal operation would be disturbing. This loss in the resnatron tube circuit is no doubt necessary to allow sufficient resistance to protect the valuable resnatron tubes. Is there any hope of reducing the loss level for this type of equipment in the future?

It is interesting to know the bushing insulation class (46 kv) used by a laboratory group for this type of d-c apparatus. The normal withstand of a 46-kv class bushing is 120 kv rms, 60-cycle, or 170-kv crest. If one arbitrarily assumes the d-c rating of the bushing comparable to the crest value, there is only a 24-per-cent margin of the withstand value over the no-load direct voltage of 137,000. I would consider this to be too little margin and would rather use about 50-per-cent margin. Since bushing insulation classes are not established considering usage on direct voltages, it will be interesting also to know if any difficulties develop in using the bushing.

In connection with the section on "Test Results," I would be interested in knowing the ripple voltage which was observed and at what value of reduced voltage output. In this same connection, how was the figure of 80,000 volts ripple extrapolated?

It would be interesting if the author gave some details of the filter choke L_1.

My interest in this paper arises from being connected with the design of a 100-kv 3.5-ampere d-c power supply which my company recently furnished for a similar application.

J. J. Halloran (Electro Engineering Works, Oakland, Calif.): There is an interesting point in this article which, from the transformer manufacturer's point of view, is worthy of further discussion; this was the report of initial tests in which ripple voltages were observed equivalent to 80,000 volts peak to peak at full output as compared to a calculated and design value of 17,000 volts.

If this condition had been permitted to continue it would not only have stressed the filter circuit components but also the transformer insulation. The fact that the authors analyzed their circuit conditions and were able to correct the resonant condition which caused this excessive ripple is extremely gratifying to the transformer manufacturer whose equipment might otherwise have been damaged.

As an example of trouble that may be experienced, this manufacturer constructed a 200-kva 500,000-volt oil-insulated transformer for half-wave rectification which, after several months in service, had insulation breakdown on the end coil, although the insulation was actually adequate to withstand full operating voltage in air and had surge shielding adequate to distribute and withstand the impulse voltages required for electric power and distribution transformers. This failure was caused by repeated arcing within the rectifier tube connected to the high-voltage terminal. Once this tube problem was recognized and corrected there was no further trouble.

It is believed timely to point out that the American Standards Association Standards *C57*[1] for transformers, regulators, and reactors cover mercury-arc tube rectifier transformers but do not pursue the subject into the field exemplified by the resnatron plate supply. It is believed that this field is sufficiently mature to warrant a more comprehensive standard, with proper guidance being given to the user of the equipment as well as to the manufacturer. Also, simple little questions like the proper marking (*X* or *H*) of a low-voltage winding of a filament transformer connected to a circuit of higher voltage than the high-voltage winding cannot seem to be resolved by the present standards to the satisfaction of all parties concerned.

It is hoped that the article will accelerate recognition of the need for adequate standards on, as the International Electrotechnical Commission so aptly puts it, "power converter equipment other than rotary converters and motor-generator sets." We are sure that such a program will receive strong support from the manufacturing industry.

REFERENCE

1. TRANSFORMERS, REGULATORS, AND REACTORS. *ASA C57*, American Standards Association (New York, N. Y.), 1948.

R. P. Featherstone, C. Ling, and **P. A. Cartwright:** In answer to Mr. Schlawin's comment on the 5-per-cent power loss in the resnatron anode protective resistors, there may well be hope of reducing this loss considerably in the future. The resistor value was specified by the resnatron manufacturer, and presumably is a conservative choice because the resnatrons are experimental units being developed on special order for this project and there is no operating experience with such vacuum tubes as yet. After the resnatrons are put into service and we have accumulated a reasonable amount of experience with them, we will be better able to judge the importance of these resistors and perhaps to experiment with changing their values.

Regarding Mr. Schlawin's remarks on the bushing rating, it is true that occasionally for short periods the system may operate unloaded at voltages as high as 137,000 volts, but normal operating procedure will not permit output voltages above 120,000 volts. Thus except for periods of the order of operator response time the margin between the d-c operating voltage and the crest of the a-c rated withstand voltage will be about 40 per cent. It is probably true, however, that if this apparatus were being designed for duty in which continuity of service is paramount a more conservative design would be in order. In our case a flashover a month or even one a week would not be significant compared to the expected rate of arc overs in the resnatrons.

So far there has been no trouble with the bushing or with the 46-kv class station post insulators used elsewhere in the system. We expect to have a low insulator flashover rate because of the favorable conditions of indoor, laboratory operation.

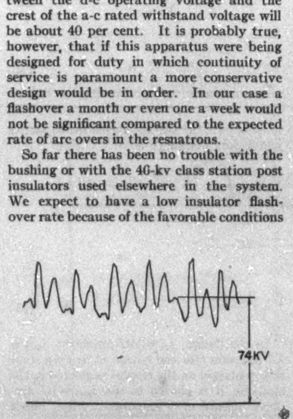


Figure 1. Typical voltage wave form at input to filter, reduced voltage, C_1 approximately 0.0013 microfarad

In answer to the question on ripple voltage, Figure 1 of the discussion may be of interest. It is a tracing of a typical oscillogram of rectifier output voltage wave form (with zero-voltage baseline) with nothing but cable and stray capacity—a total of about 0.0013 microfarad—in the position of C_1, Figure 1 of the paper. This particular oscillogram was taken with 1/60 second sweep, and rectifier output 74,000 volts at 2.7 amperes. The peak-to-peak ripple voltage is approximately 47,000 volts, corresponding to 70,000 volts ripple at full output voltage on a linear extrapolation.

The relative size of the ripple was observed to vary under different load conditions, but no study of its dependence on load impedance has been made. The worst case observed occurred at 50,000 volts and 1.63 amperes rectifier output, with 41,000 volts peak-to-peak ripple. This corresponds to 90,000 volts peak-to-peak ripple at full output voltage with linear extrapolation. In this case, however, the ripple was probably increased somewhat by a 6-per-cent unbalance in the 3-phase primary voltages which was present at the time of that particular test.

It is interesting to note that the presumed shock excitation seems to occur at something over twice the natural 360-cycles-per-second ripple frequency of the system, varying somewhat from phase to phase, presumably with the value of leakage reactance of the three transformers. In the scale model tests it was possible to demonstrate generally similar output voltage wave forms, the frequency of the superimposed shock excitation depending on the adjustment of the transformer leakage reactance and on the capacitance placed across the input to the filter.

Mr. Schlawin has asked about the design of choke L_1. The calculations in the Appendix establish its minimum value as 10 henrys. The cost of a 10-henry 4-ampere choke insulated for the full supply voltage end-to-end would be quite high, and since this system will always be run up slowly we do not require a choke of full voltage rating. Therefore L_1 was specified to withstand 20,000 volts rms end-to-end; this was expected to take care of normal ripple and the load voltage swings, with an adequate margin.

The case of L_1 is oil-filled and air-cooled. The coil is made in two pies connected in series, the beginning of each connected to core and case. It is wound with number 13 paper-enamel wire, with 0.01-inch 100-per-cent rag insulating paper between windings. The core is made up of four Hipersil loops, each $3^{1}/_{8}$ inches by $2^{5}/_{16}$ inches in cross section. Maximum flux density (d-c plus a-c) is 16,000 gausses. It is a linear reactor, most of the reluctance being in the air gap. Its test inductance is 16.3 henrys, its cold d-c resistance is 16.6 ohms. It is protected by a spark gap, which is normally set to break down at 28,000 volts peak. Our attention was first drawn to the problem of abnormal ripple voltage by the firing of this spark gap as the voltage was raised under load.

The authors agree with Mr. Halloran that there is a definite need for improved standards on electronic power converters.

A Nonarmored Submarine Telegraph Cable

C. S. LAWTON
MEMBER AIEE

L. H. HUTCHINS, JR.
MEMBER AIEE

Synopsis: Most of the world's submarine cable mileage lies in soft ocean bottom where the customary armor wires are not required for protection of insulation against abrasion. An experimental length of unarmored cable having polyethylene insulation over a central conducting core of cadmium bronze sufficiently strong to permit of laying and repair has recently been laid. This paper describes the mechanical and electrical design features, the problems encountered in laying and repair at sea and their solutions.

THE IDEA of omitting the steel armor wires of deep-sea cables has been in the minds of ocean-cable engineers for a good many years. Necessarily, a cable must be of such construction mechanically that it can be laid on the ocean bottom in sound condition. But once there, a cable lying in the silt of which the ocean floor is composed at depths of the order of one thousand fathoms is subject, under normal conditions, to virtually no physical disturbance, and hence may be presumed to require no extraordinary mechanical protection. It is only the physical stresses imposed on cable during the laying and repairing processes which have made it necessary to include a heavy, expensive steel sheathing in the construction of deep-sea cables. Therefore, engineers have reasoned, cable companies pay a sizable proportion of their deep-water-cable costs solely to put cables on the bottom, not to keep them there. One logical approach, then, to the ever-present problem of reducing cable costs is quite conceivably to be found in the omission of the sheathing. Unfortunately, however, the lack of appropriate materials hitherto has made it impossible to reconcile essential physical and electrical requirements in such a way as to construct a practical, economical cable.

A submarine telegraph cable for deep water service must have several diverse properties, as follows:

1. It must have sufficient tensile strength to support the dead weight in sea water of the length of cable between the ocean bottom and the laying ship, plus the live load caused by the vertical movement of the ship on the ocean swell, at depths that might vary from 1/2 or less to $2^1/_2$ nautical miles, and sometimes more, as the laying operation proceeds. The necessary relation between the strength of the cable and its weight in sea water is expressed as the so-called "cable modulus," which is defined as the ratio of the breaking strength of the cable to the weight in sea water of a nautical mile of the cable. The value of this modulus thus represents the length of cable in nautical miles which, when suspended vertically in sea water, is self-supporting. The required minimum value for modulus will vary with the maximum depth in which a cable will be laid. The safety factor will be given by dividing the modulus by the depth.

2. It must possess sufficient radial compressive strength to transfer the load represented by the weight of cable below the ship onto the drum and sheave over which it is payed out, without mechanical failure or permanent distortion of its component parts. That property which is the measure of ability to resist such forces is commonly called, when applied to a single material, the modulus of compression. For a composite structure like an armored cable, however, the usual definition of modulus is not applicable, and consequently such an attribute has never been quantitatively assigned thereto. For the relatively simple structure of a nonarmored cable, the value of modulus possessed by the outer covering is of great importance and must be considered, since the entire tensile load is being carried by the central member of the cable.

3. It must be of such construction mechanically that sudden changes of tension thereon do not produce distortions such as kinking and knuckling. The former is evidenced by the formation of small loops along the cable as a result of severe twisting about itself, and the latter by the ejection of the conductor from the insulating envelope as a result of the formation of a sharp bend in the conductor. Both of these phenomena result in cable failure: a bad kink will usually cause the conductor to part when tension is regained, and a knuckle ruptures the insulation immediately to produce a fault to ground.

4. It must provide all the electrical properties essential to its operation as a transmission line for telegraphic communication. The so-called "speed of signalling" of a telegraph cable, that is, the maximum rate at which character impulses can be transmitted with sufficient intelligibility, depends primarily upon the capacitance of the insulation and the resistance of the conductor. It is the product of these two parameters that is known as the "time constant" of an electric circuit, and, in earlier days, when applied to cables, was called the "KR Law," which stated simply that the signalling speed was inversely proportional to the product of d-c resistance and inductive capacitance. So propounded, it neglected skin effect, inductance, and conductance, which, at the frequencies of transmission then employed, was a good enough approximation; and, as a matter of fact, still is. Therefore, any appreciable length of cable spliced into an existing line must possess a unit KR value no greater than that of the line, if it is not to slow down the cable circuit as a whole, that is, produce a decrease in the signalling speed employed.

The insulating material of a cable also possesses an electrical parameter that has no effect on signalling speed unless its value is so small as to cause the conductance, ordinarily insignificant, to become large—dielectric resistance. It is, nonetheless, a very important property, for it is used as a measure of insulation quality throughout the life of a cable. The first warning of deterioration and impending failure of the dielectric is usually given by abnormal behavior of the galvanometer in the measurement of electrification current. Such a condition can be discovered, moreover, long before it affects transmission.

5. Once laid, it must be inert in the sea water in which it is immersed. The insulating material surrounding the conductor must deteriorate neither physically nor electrically at a rate rapid enough to interfere with successful operation within a period of time that, in the ocean-cable art, is counted in decades. In deep water, a cable life of more than 50 years is not uncommon. Physical instability, in so far as it has no direct effect on the electrical characteristics and does not act to change them, can be tolerated, but only to a limited extent. For example, if a cable loses its mechanical virtues so rapidly that it cannot be raised to the surface within a reasonable time after being laid, repairs are difficult to make and a fault could result in expensive replacement. On the other hand, under no condition can electrical instability be tolerated, for so long as it exists the cable is an operational liability. If resistance and/or capacitance increase with time, compensatory changes must be made in terminal networks and amplifiers, which can become expensive. When they become impossible, replacement of the cable is the only alternative to the continuance of reliable communication.

It is apparent from a study of these requirements that the omission of the steel wires from the outside of a deep-water cable presents some interesting mechanical problems. First, it is these wires that give the cable its effective tensile strength, for they are of high-carbon relatively low-elongation steel. The conductor, the only other metallic member, is of soft-

---

Paper 53-163, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted October 22, 1952; made available for printing December 5, 1952.

C. S. LAWTON is with the Western Union Telegraph Company, New York, N. Y. L. H. HUTCHINS, JR., is with the Simplex Wire and Cable Company, Cambridge, Mass.

drawn high-elongation copper and, therefore, contributes substantially nothing thereto. Hence the task of supporting the weight of the cable in suspension falls entirely on the steel. Second, since these wires constitute a helical sheath that completely encloses the core (which is the term applied to the insulated conductor), they serve efficiently as a protective barrier against radial forces. Hence they furnish the cable with the radial compressive strength essential to its laying. The sheathing also poses a minor electrical problem, for, being metallic, the wires help to conduct the earth-return currents and thereby influence the form of the sea-return path. Thus not only circuit resistance but inductance as well is affected by their presence. A cable without this steel armor, then, must be so designed that all these problems are solved with practicality and economy; and it was the commercial production of two recently developed materials that made it possible to do so.

One obvious means of eliminating the outer sheathing wires is to combine as many of their functions as possible with that of the central conductor, and thereby cause the conducting member to serve also as the strength member. The complete cable need only consist, then, of an inner metallic material and an outer insulating material; provided both of them, individually and co-operatively, possess the characteristics imposed by the demands of the service.

With the entire responsibility for supplying the tensile strength of the cable falling on the conductor, the mechanical arrangement of the elements thereof must be given careful consideration. The simplest structure is of course a single, solid wire, but unfortunately it is not a practical one. The possibility of flaws, although remote to be sure with the present high development of the art of drawing and heat treatment, must nevertheless be considered. Also, since joints between coil lengths of bare wire could not be made by welding or brazing, because the high temperatures necessary would change the physical characteristics, and a mechanical-type joint cannot be insulated by an extrusion machine, an excessive number of finished-cable splices would be necessary.

With a solid wire ruled out by reason of mechanical limitations, it becomes necessary to design a conductor made up of several strand elements, and so disposed that the tensile strength of the total area of conducting material is most efficiently realized, with respect to its weight. The most common method of stranding is to spiral six wires, or strands, around a central one, all being of equal diameter. Such a construction, however, is very liable to place a disproportionately large share of the total load on the center strand, by reason of its position within the structure, and result in a reduction of ultimate strength by as much as two-sevenths of the available total. A stretched or ruptured center strand would cost one-seventh, and a brazed surrounding strand another seventh, assuming they occurred at the same cross section, which is quite possible. Adequate spacing of brazes in the individual strands during fabrication would have to be specified, else the conductor could be further weakened to the point of inutility. Suppose, instead, the conductor is arranged as an annulus of n strands of uniform cross section in a single layer. With this configuration, each strand is subjected to the same forces as every other strand and the minimum strength developed can be no less than $(n-1)/n$ fractional part of the possible total, and this at a point where a single-strand joint occurs. The final detail of the mechanical design then becomes the determination of a practical value for n, not a difficult decision to make. An annulus of ten wires was decided on, on the basis of:

1. Feasibility of manufacture, in that appropriate machines were available for the stranding operation.
2. Adequacy of numbers, for it was essential that a sufficient number of strands be present to cause a reasonable reduction on the total available tensile strength at points of strand brazes.
3. Ease of splicing, because there would not be so many strands that complete-conductor (or finished-cable) splices would be exceptionally difficult to make.

With a fundamental configuration decided on, the next step in the design of the nonarmored cable was to make a complete electromechanical analysis of all the parameters of the conductor and dielectric affecting the performance of the cable. Once those were established, it was only necessary to decide on the values of speed and modulus that were considered essential for the cable to possess. A speed constant was selected comparable to that of existing transatlantic telegraph cables. In the selection of the value of modulus, however, present practice could not be used as a criterion. The modulus of armored cables usually runs above 7 nautical miles, not because they are designed to be laid in such extreme depths, but because the armor wires, being exposed to the ravages of the salt water, corrode in the course of time and consequently lose their strength. Therefore, if a ship is to be able to raise the cable to the surface several years after it is laid, within its normal life span, in order to repair a fault, allowance must be made for this diminution in modulus by making it larger than necessary when new. A nonarmored cable, on the other hand, has its strength member within the insulating sheath, completely protected from the action of the sea. Hence, no loss in modulus need be designed for, and the initial value can be as small as is adequate to provide for a reasonable factor of safety during laying and/or recovery. Since deep-sea cables are rarely laid in depths exceeding $2^{1}/_{2}$ nautical miles in the north Atlantic, that value appeared to be a safe maximum to use for the dead weight. More or less arbitrarily, a figure of two for the combined live weight and safety factor was selected, producing a value of five for the modulus.

The one remaining problem, now, in the complete design of the unarmored cable is the determination of the proper materials for the conductor and dielectric. The material selected for the conductor was cadmium-copper, 1 per cent by weight cadmium and 99 per cent copper, having a tensile strength of 90,000 pounds per square inch, a conductivity of 80 per cent International Annealed Copper Standard, and a specific gravity of 8.89 based on pure water. Not only does cadmium-copper (also called cadmium-bronze) possess a remarkable combination of strength and conductivity but it also

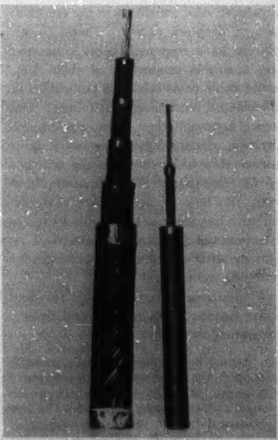


Figure 1. Armored deep-sea cable (left) and nonarmored cable (right)

possesses another extremely useful property—high fatigue resistance—which imparts to it exceptional torsional strength, or the ability to withstand considerable twisting and bending without failure.

The essential properties of the insulating material are its specific gravity and dielectric constant. Actually, not much choice is presented in the selection of a suitable insulator, for only one material, polyethylene, is currently available that makes it possible to construct a cable that is at once technically and economically practical. With a specific gravity of 0.925 and a dielectric constant of 2.28, polyethylene meets all other physical and electrical requirements. It has good compressive strength, good abrasion resistance in heavier walls, fair resistance to cold flow compared with other types of submarine insulations, and an exceptionally low rate of water absorption. It possesses an amazingly high resistance constant, high dielectric strength to both alternating and direct voltages, and low a-c conductance or loss factor, not in themselves of first order importance, but very useful attributes.

To test the new design under service conditions, Western Union decided to construct about 20 miles for insertion in a working cable. Although designed for transatlantic use, one of their coastwise links was chosen for the test because of its easy accessibility, and because a landline fall-back facility was readily available in case of trouble. The first nonarmored cable was manufactured in May of 1950, and had the following construction and constants:

1. 17 nautical miles—type *D-485/600* bronze and polyethylene nonarmored deep-sea ocean telegraph cable.

2. Conductor number 6 American Wire Gauge (AWG) (approximate)—ten number 16 AWG (0.051 inch) hitenso *BB-961* hard cadmium-bronze wires around a polyethylene thread of 0.128-inch diameter, 3.00-inch left-hand lay (preformed and postformed).

3. Wall, 0.180 inch = 0.580 inch insulating diameter polyethylene.

4. Cable modulus = 5.05 nautical miles.

5. Speed constant = 0.69 ohm-microfarad per nautical mile of cable length.

Figure 1 illustrates the physical comparison between the new cable and a conventional transatlantic-type armored cable embodying standard practices.

Three fundamental problems arose in connection with the proposal to repair breaks in existing cable lines by the insertion of a nonarmored type of cable, as shown by Table I which gives a comparison between typical conventional armored deep-sea telegraph cables and the nonarmored design. The problems arising were:

Table I. Comparison Between Conventional and Nonarmored Telegraph Cable

	Armored		
	Coastwise	Trans-Atlantic	Nonarmored
Diameter, inches	0.90	1.04	0.58
Weight in sea water, pounds per nautical mile	1,950	2,690	360
Breaking strength, pounds	17,250	20,400	1,800

1. Handling nonarmored cable with ship's machinery designed for conventional types is like trying to play a trout with gear designed for tuna. The pull necessary to revolve the cable drum with driving gears de-clutched and brakes lifted is 600 or 700 pounds and the usual ship's dynamometer is completely insensitive to tensions below 1,500 pounds. Repair ships like the *Lord Kelvin* are equipped to pay out cable over the bow, but the danger of such light cable fouling the ship's propellers (twin screws are standard in cable ships for maneuverability) obviously made it impractical to pay out the nonarmored cable over the bow.

2. The new cable, having a breaking strength far below that of steel-armored cables, does not possess sufficient tensile strength to support, in a depth of water for which it is designed, the weight of such a length of armored cable. Hence, how may this light, relatively weak cable be spliced to the heavier one, and the combination successfully lowered to the ocean floor?

3. The new cable, having a specific gravity in sea water considerably below that of old cable, will sink far more slowly. Hence, at what laying speed will it be necessary for the ship to run, in order not to trail the cable too far behind it?

As a solution of the first problem, a light, easily-portable paying-out machine was constructed and installed aft, together with a more sensitive brake and dynamometer. A paying-out drum of 39 inches in diameter, equipped with a fleeting knife, was used. It could have been even smaller. The width of tread was 6 inches, which was ample. A duplicate of this steel casting was mounted on the same shaft to serve as a brake drum, fitted with an external steel band having a molded segmented lining. Water cooling was avoided by the simple expedient of using a brake of ample area to dissipate the heat generated. The brake lining provided a frictional coefficient that remained substantially constant at all speeds, including the static condition. A lever arm was provided for loading the brake with weights. The frictional drag of the cable coming up from the ship's tank through bell-mouths and around roller leads was found to be enough to ensure tightness of the turns on the paying-out drum, and although means were provided for putting additional drag on the cable leading onto the drum, they were not used. From the drum, the cable passed through a dynamometer of conventional cable-ship type, but constructed to a much smaller scale, so as to be correspondingly more sensitive. Leaving the dynamometer, the cable left the ship by way of a stern sheave, also 39 inches in diameter, mounted directly abaft the paying-out gear and overhanging the after rail.

Coming to the second problem, consideration was given to buoying the first splice but this would have meant leaving slack in the cable when the buoy subsequently was removed and the bight of cable was dropped. While not serious for an experimental job, this would be unacceptable as a regular practice, both transmission-wise and economically. Therefore, it was decided to interpose between the steel-armored and the nonarmored cables two intermediate designs that would make the transition less abrupt by providing a tapering of strength and weight along the section. The first of these was constructed as follows:

1. One-half nautical mile (3,044 feet)—type *D-485/600-24/13* cadmium-copper and aluminum-armored deep-sea cable.

Over finished nonarmored cable apply:

One 70-pound cutched jute serve, applied wet

Twenty-four number 13 British Wire Gauge (BWG) (0.095 inch) type *56S-H* aluminum-alloy wires (preformed)

One 17/3 impregnated jute serve

Over-all diameter = 0.990 inch.

and then spliced to the second:

2. One and one-half nautical miles (9,131 feet)—type *D-250/430/600* copper, steel, and polyethylene nonarmored transition cable.

Conductor number 6 AWG (approximate)—One strand, 0.117-inch not-tinned copper; Ten strands, 0.051-inch improved plow steel around; 3.00-inch left-hand lay (preformed and postformed)

0.180-inch wall = 0.580-inch diameter polyethylene

which was in turn joined to the 485/600 nonarmored cable at the factory prior to shipment. Table II gives a comparison of the four types involved in this experiment.

The moduli between cables become 5.2, 4.3, and 3.2 nautical miles in passing from steel-armored cable through the transitions to the nonarmored cable, which

Table II

Cable	Breaking Strength, Pounds	Weight in Sea Water, Pounds per Nautical Mile	Modulus, Nautical Miles
Steel armored	17,250	1,950	8.8
Aluminum armored	10,000	1,150	8.7
Copper and steel	5,000	560	8.9
Cadmium-bronze	1,820	360	5.0

proved to be satisfactory for the initial installation in 1,200 fathoms of water.

The solution of the third problem involved first the calculation of the rate at which this new cable would sink into the sea, and second the determination of the relationship between the angle of inclination of the cable with the water surface and the speed of the laying ship. In both cases, only theoretical studies were made, for the equipment necessary for performing the proper experiments was not available. Moreover, precise results were not needed, but merely an idea of the order of magnitudes involved. These calculations showed that pay-out could probably be made at a speed not too much slower than normal, with a reasonable laying angle prevailing.

Echo sounding showed the average depth along the proposed route to be about 1,250 fathoms. Sampling indicated soft mud bottom.

In order to make possible the paying-out of the transition pieces over the ship's bow and the nonarmored cable over the stern, special preparations had to be made. These consisted of leading the end of the first transition piece up from the ship's tank, through the after paying-out gear and dynamometer, thence over the stern sheave, and forward around the starboard side of the ship outside everything (even the ship's rigging) and in again through the bow sheave frame. Via this route, both transition pieces and a short length of nonarmored cable were transferred to the ship's forward well deck, where they were coiled down. The next step was to reverse the coil in order to bring the end of the first transition piece to the top, ready for splicing to the cable in circuit. Meantime, the bight of nonarmored cable was secured along the outside of the ship with rope yarn lashings which could be cut away quickly.

The working cable was hooked, cut, and raised on a grapnel on the east end of the proposed insertion, tested through to Canso, Nova Scotia, suitably prepared for splicing, and temporarily buoyed off. No attempt was made to recover the cable lying on the bottom over the distance within which it was to be replaced, the ship proceeding directly to cut and raise the cable to the westward near the position for splicing on and commencing the pay-out.

The suspended end was brought inboard over the center bow sheave and the starboard picking-up gear and led to the splicing area. After testing through to Hammel, Long Island, the first ship splice (steel-armored to aluminum-armored cable) was made in the conventional manner, by overlapping the steel sheathing with the aluminum sheathing.

The pay-out was begun under a tension of 23 hundredweight, with the first splice passing over the bow without incident. As more of the aluminum cable went out, tension was gradually decreased so that by the time the splice to the second transition piece went over the bow, the tension was only 16 hundredweight. From there on, the tension was tapered off until the splice to the regular nonarmored cable went over. As the cable in suspension gradually became lighter, the angle at which it entered the water diminished, bringing the point of entry farther and farther aft along the ship's starboard side. In order to reach the very low tensions required, it was necessary to keep the paying-out drum in gear and drive it with the engine rather than depend upon the brake. Adjustment of tension near the tail end of the aluminum piece was a matter of judgment as the dynamometer came to rest against its bottom stop before the splice was reached.

As the last of the regular nonarmored cable coiled forward left the deck, the ship was stopped and the cable was held on a manila stopper. The nonarmored cable was lifted off the paying-out drum and dynamometer, and as the stopper rope was eased away, the bight passed through the bow-sheave frame. The lashings along the side were cut one by one as tension came on them until, with the cable suspended freely from the stern sheave, the manila stopper rope was cut away and the pay-out was resumed from aft.

Tension was maintained in the region of 420 pounds, which was about 7 per cent less than the vertically-suspended weight at a depth of 1,250 fathoms in order to avoid residual cable tension on the bottom. A paying-out speed of 4 knots was found to be about the practical limit because the cable was emerging from a small after tank, and the particular lead employed tended to produce some slight whipping.

At this speed, the cable entered the water at an angle of approximately 10 degrees with the horizontal. The buoyed east end was reached with about a mile and a half of cable remaining in the tank. The buoy had been set to the westward by the current, so that in picking up the moorings, the ship had to move eastward, thus making it necessary to continue paying-out the nonarmored cable. By the time the end of the steel-armored cable has been raised to the bow on the buoy moorings, all but 200 fathoms of the nonarmored cable had been paid out over the stern.

With the cable held on a stopper, the surplus was cut off and the end was then transferred to the bow around the outside of the ship for final testing and splicing. The final splice was made by first serving the last few fathoms of nonarmored cable with cutched jute, then carrying the steel sheathing wires of the armored cable well beyond the core joint. The splice so made was then served over with tarred hemp spun yarn in conventional fashion, the bight was lowered over the bow on manila rope and the latter cut away, thus releasing the bight of cable and completing the laying operation.

◆

Discussion

J. J. Gilbert (Bell Telephone Labs., Inc., New York, N. Y.): The nonarmored cable represents a notable departure from submarine cable conventions and the authors are to be congratulated for their ingenuity and courage in engineering the project that they describe. Steel armor has served a useful function through the years, but at a price. The conventional requirement is that the armored cable shall be strong enough to support itself in a depth about three times that in which it is to be laid. To meet this requirement in the case of a deep water cable it is necessary to assign a large part of the strength of the armor wires to the support of the armor itself so that the support furnished to the remainder of the cable is marginal and a lot of steel is needed. This is not only a hardship from the cost standpoint but the tensions involved in laying and recovering such heavy structures are not healthy for the electrical heart of the cable.

The nonarmored cable described in the paper seems to offer a happy solution to these problems. The fact that the strength member is put in a position where it is not subject to corrosion and abrasion is properly emphasized by the authors as one of the advantages of the new structure. This should permit a reduction in the conventional safety factor but the extent of this gain can presumably only be determined by further trials in deep water. From the standpoint of life, conditions certainly look favorable for the nonarmored cable. Time, of course is needed to confirm this. The present trial will be watched with great interest by the submarine cable community.

There is another aspect of the nonarmored cable that may become of importance in cases where high speed operation is employed. Theory indicates that at the higher frequencies in the telegraph range the return conductor of the conventional cable, comprising the armor wires and sea water in parallel, has a higher resistance and a lower inductance than if the armor wires were absent. At high frequencies the nonarmored cable should therefore have superior transmission characteristics.

I. S. Coggeshall (Western Union Telegraph Company, New York, N. Y.): The insertion of a nonarmored sample within a section of armored cable was a necessary step despite its complications, it being evident that the laying problems involved would have to be met whether armorless cable is to be employed in the future either to replace old cable or to be laid as new between shore approaches. Attention will now undoubtedly be focussed on its usefulness for new, deep-sea cables of single conductor or coaxial types.

The exact nature of the sea bottom is surmised to be suitable for the laying of long stretches of unarmored cable in ooze; but there is just enough of an element of doubt as to the existence of rock outcroppings and suspension-fraught gradients to make the determination to lay a transatlantic unarmored cable, at a cost of several million dollars, a matter of courage and faith as well as of skill in design and manipulation.

Perhaps the author would care to comment on these features, as well as on the question regarding the vulnerability of unprotected polyethylene to the ravages of the teredo worm.

Kenneth S. Wyatt (Phelps Dodge Copper Products, Yonkers, New York): As a newcomer to the submarine communication cable field, I should hesitate to enter into this discussion were it not for long experience with polyethylene, which made this new cable possible. The complete elimination of armor represents a sharp break with tradition, and I can well appreciate the courage of the authors in this bold move. They have backed their judgment with installations of considerable magnitude (about 100 miles). It is a real joy to see such independent engineering thinking and action.

The new conception and the reasoning leading up to it are presented in a logical and clear-cut manner. Especially noteworthy is the ingenious method of avoiding a sharp transition from a stiff rigid cable to a light flexible one by tapering the stiffness with intermediate types of cables.

The properties of polyethylene must be carefully examined to see if it is capable of withstanding the considerable strains to which a deep-sea cable must be subjected. Especially must we consider cold flow, which would permit decentralization of the conductor, abrasion resistance, and compressive strength. In 1945 I introduced polyethylene as a sheathing material for extra high voltage cables of the pipe-line type. The polyethylene sheathed single cores of several thousand feet length are wound up on reels sometimes weighing over 15 tons. The pressure on the polyethylene sheath of the inner turns is very high, yet no flow of the polyethylene has been noticed, nor has there been any trouble with between 100 and 200 miles of such cable.

Our Habirshaw Research Laboratories have done a lot of experimental work on polyethylene. With regard to the possibility that the conductor of Mr. Lawton's cable might cut through the insulation as it passes over the shieve, the following test might be cited: a cable having a conductor of 0.125-inch diameter insulated with polyethylene to a diameter of 0.46 inch was bent over a mandrel 6.3-inch diameter, and a 63-pound weight hung on each end, giving a pressure on the insulation of 160 pounds per square inch. After $7^1/_2$ hours at 98 degrees centigrade, no decentralization could be observed. In the case of the new submarine cable the pull is 420 pounds, giving a vertical component of only 73 pounds. This is far too low to cause cutting or cold flow especially at ambient air temperatures.

Polyethylene has excellent abrasion resistance because it is self-lubricating. Short lengths of heavy power cable, one with lead sheath, and others with a polyethylene jacket over the lead, have been dragged down a concrete road at 3 miles per hour. After 1/2 mile the lead sheath is in bad condition but the polyethylene is merely scuffed. However, polyethylene slices rather easily.

The pressure at a depth of 1,250 fathoms approximates 3,400 pounds per square inch, the temperature being perhaps 40 degrees Fahrenheit. It is difficult to obtain exact figures for the compressive strength of polyethylene, but we believe 3,000 to 4,000 psi at 20 degrees centigrade to be reasonable. At 1,250 fathoms it seems the pressures are approaching the ultimate compressive strength of polyethylene, especially as a certain amount of shear is involved.

For very great pressures we have been turning to a much harder material without any cold flow in the magnitude range under discussion, namely, flexible polystyrene (styroflex).

The new submarine cable has great promise for areas which are free from disturbances and where the bottom is smooth. It is an event of considerable economic importance.

C. S. Lawton and **L. H. Hutchins, Jr.**: A natural question has been raised as to the ability of unprotected polyethylene to withstand the ravages of the teredo worm which has caused a great deal of damage to unprotected gutta percha insulation in the older cables, particularly in the more temperate zones. Frankly we do not know the answer. It is one of the things we are hoping to learn from this experiment.

However, two circumstances seem to justify a reasonable doubt that this cable will be attacked. The surface is so smooth and waxy that a worm would have a hard time getting started, and experiments have shown that materials susceptible to attack are more apt to be damaged when the worm can get in between two surfaces against which it can brace itself. (This, of course, is the case with conventional cable when there are openings in a sheath of armor wires separated from the insulation by a soft jute bedding.)

The other circumstance is that we have no proof that boring worms live in the greater depths of the ocean. Western Union's maintenance records do not disclose any instance of teredo attack in depths exceeding 1,000 fathoms, although covering thousands of miles of cable exposed for many years at these depths, much of which has no brass tape protection on the gutta percha. Adaptation of a boring organism to a habitat of practically nothing but soft mud and ooze, although possibly achievable through natural processes, presupposes a motive. If we assume that a motive exists, though obscure, what happens to the boring function with nothing harder than mud to work on? It seems likely that the organism itself will cease boring or move to another location. But this is only speculation and as such could turn out to be quite wrong.

Perhaps the best way of summarizing the matter would be to say that there was more to be learned by omitting teredo protection than by providing it on this experimental length.

With reference to Mr. Wyatt's discussion, it should be pointed out that the only installation to date is somewhat less than 20 miles in length. We wish it could have been 100.

The most common bottom temperatures at depths exceeding 1000 fathoms in the North Atlantic are between 35 and 40 degrees Fahrenheit the year around. Under these conditions much greater resistance to abrasion is afforded by polyethylene than at higher temperatures. Notwithstanding that the nonarmored cable compares favorably with armored cable in resistance to abrasion under service temperatures and where bearing pressures are proportional to weights in water, it is not considered suitable for use where there are bottom currents such as exist in shallow water. Currents which would not move a heavier cable at all could cause movement in a nonarmored cable which sooner or later would lead to failure, either through chafing of the insulation or vibratory fatigue of the conductor, although as already noted, cadmium copper has excellent resistance to fatigue failure.

As far as the effect of hydrostatic pressure is concerned, there is no difference between the force exerted on conventional cable core and nonarmored cable. The force being equal all around the periphery, there should be no tendency to cause displacement of the conductor. Polyethylene-insulated cable has been successfully laid in nearly 3,000 fathoms without developing any hydrostatically induced weakness.

Of far greater concern is the tendency of such cable with conventional armor to fail through knuckling of the conductor upon the release of the high tensile and torsional stresses imposed upon it during laying and repairing operations. Ingenious means have been adopted for overcoming this tendency which give promise of being successful but the fact remains that the weight of the armor is a great handicap: so much so that it rarely, if ever, has been possible to pick up conventional cable from depths exceeding 1,000 fathoms in fit condition to re-lay in deep water, even when it has suffered little or no deterioration from exposure to the sea water. That these defects are not caused by hydrostatic pressure is obvious because they can be duplicated readily by tests on shore under atmospheric pressure.

Not the least of the advantages of nonarmored cable lies in the almost complete elimination of torsional forces because of the close grouping of the helically shaped strength members about the neutral axis, and the fact that most of the elongation of the conductor which occurs upon the application of tension is achieved through distortion of the conductor helix, rather than through any reduction in cross-sectional area of the individual wire strands. (It is the stressing of soft copper beyond its low yield point in conventional cable which causes the knuckling tendency when the polyethylene pulls back strongly upon release of tension.)

No obstacle is seen to the design of nonarmored coaxial cable, but as this type usually is associated with submerged repeaters, reinforcement to provide enough strength to carry them to the bottom and raise them for servicing will be required for a reasonable distance on each side of each repeater, except possibly in the case of articulated repeaters of very light weight, and the economic advantage of the nonarmored design will depend upon the spacing of repeaters and their weight in sea water.

That this experiment has aroused so much interest is not too surprising. At a time when there are so many indications of a genuine resurgence of interest in ocean cables as the oldest and most dependable medium for international communications, anything which may affect their cost, their longevity and ease of repair at great depths is bound to be of real concern to those of us in the communications field who are aware of the spectacular advances which are now being made in transmission technique.

There is every reason to believe that this cable can be recovered and re-laid without impairment of its strength or coiling properties; and after a test of about 3 years (it has been in circuit almost 17 months now), assuming it still to be electrically sound, it is planned to do just that, utilizing the opportunity to examine every inch of it for evidence of teredo attack, chafe, and so forth, of course, and relocating it in another area where conditions are known to be none too good.

It is hoped that it will be possible to submit a further report at that time.

Accurate Radio-Frequency Microvoltages

MYRON C. SELBY
NONMEMBER AIEE

Synopsis: The questionable accuracy of radio-frequency microvoltages has been of great concern to the radio field for many years. There is an urgent need for a simple, yet reliable, source of microvolts for measurements in general and for radio receiver sensitivity determinations in particular. Extremely simple devices, which seem to satisfy that need most adequately, were recently developed. They are sources of potential drop obtained across a known resistance through which known currents flow. These devices provide constant voltage sources of accurate microvolts over a range of 1 to 10^5 and wider, at all frequencies to 300 megacycles and higher. They are adaptable for balanced as well as for unbalanced sources. Their electrical constants are simply determined by using known direct voltages and currents. Basic principles, design features, and applications are discussed.

THE RADIO and electronics field has been facing the problem of accuracy of radio-frequency (r-f) microvoltages since sensitivity of radio receivers came into prominence as a competitive index of performance. The reasons for the continued existence of this problem are too well known. There is no need to discuss them other than perhaps to indicate the major ones: namely, 1. the uncertainty in voltage source and load (receiver input) impedance values, 2. the uncertainty in performance of attenuators over wide frequency ranges and attenuations of up to 100 decibels and in many cases of higher values, and 3. the uncertainty in the accuracy and stability of devices monitoring the input voltages to the attenuators.

Considerable progress was apparently made as time went on, and a large section of the field has managed to attain high accuracies despite these difficulties. A recent poll of the field, instigated by the National Bureau of Standards in 1951 to determine the urgency of this problem, revealed the following status. Of about 70 standard voltage-generator manufacturing and applying laboratories (six of them in United Kingdom and France) about one-half believed they had available 1 microvolt at frequencies to 1,000 megacycles with an absolute accuracy of 30 per cent or better. The indicated accuracy was naturally better at lower frequencies. The other half was uncertain of their values in various degrees, in many cases exceeding 100 and in some 200 per cent. About one-half of them desired absolute accuracies of 0.5 to 6 per cent for all frequencies to 1,000 megacycles; the rest desired accuracies of 50 per cent or better.

The critical need of reliable tools to supply or measure microvolts accurately was therefore still very much in evidence. Devices apparently fulfilling this need most satisfactorily were recently developed. They seem to meet the most persistently demanded feature of a standard of microvolts, namely, extreme simplicity and reliability. These devices, referred to, for want of a better name, as "Micropotentiometers,"[1] are described later.

General Description of the Micropotentiometer

Basic Requirements and Principle Operation

In searching for a source of accurate and reliable r-f microvolts the following basic requirements seemed indispensable or highly desirable:

1. The output voltages of the source had to be known irrespective of loading conditions, that is, a constant voltage source was necessary.
2. Freedom from frequency corrections was essential at least over reasonable frequency ranges.
3. A reliable, simple, and rugged physical construction with a very minimum of component parts was most desirable.
4. The simplest and a very minimum of calibration requirements were essential.

All of these requirements seem to be satisfactorily achieved with the Micropotentiometer. In basic principle it is a source of potential drops obtained across a known resistance through which known currents flow. These resistances are of the order of milliohms (for microvolt levels), and therefore constitute an essentially zero source impedance (constant voltage source) for all practical present-day needs. A cardinal requirement of

---

Paper 53-22, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19-23, 1953. Manuscript submitted August 29, 1952; made available for printing November 10, 1952.

Myron C. Selby is with the National Bureau of Standards, Washington, D. C.

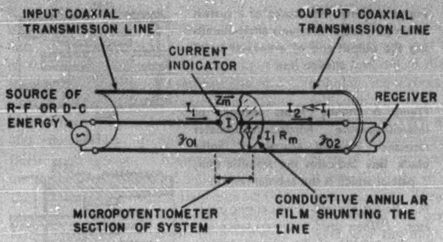
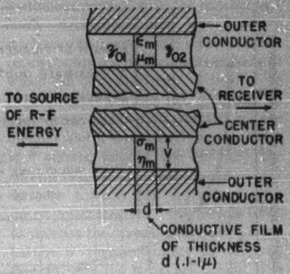

INPUT COAXIAL TRANSMISSION LINE
CURRENT INDICATOR
OUTPUT COAXIAL TRANSMISSION LINE
SOURCE OF R-F OR D-C ENERGY
RECEIVER
Z_m
I_1
I
$I_2 << I_1$
$I_1 R_m$
γ_{01}
γ_{02}
MICROPOTENTIOMETER SECTION OF SYSTEM
CONDUCTIVE ANNULAR FILM SHUNTING THE LINE

Figure 1 (left). Equivalent circuit diagram of system employing a Micropotentiometer

Figure 2 (right). Section of Micropotentiometer showing position of annular element in a low dielectric coaxial line



this impedance, as previously implied, was that its reactive component be negligible compared with its resistive component. This suggested the annular type of resistance element construction. The equivalent circuit of the Micropotentiometer together with the input and output circuits is that of a coaxial line shunted by a conductive film in a plane normal to the axis of the line, as shown in Figure 1. A source of r-f energy supplies current to the Micropotentiometer section of the system through a coaxial line and the voltage output V, from the Micropotentiometer, available across the annulus, is fed to the receiver either directly or through an output line. To a very good approximation the output voltage V is simply the product of the input line current entering the annulus and the d-c resistance of the annulus.

To show that this is true the transfer impedance of a thin conducting film placed as already indicated (see Figure 2) will be considered under the following assumptions:

1. The characteristic impedance of the coaxial lines is of practical magnitude, for example, any value between 20 and 200 ohms.
2. The bridging film has a finite high conductivity approaching that of silver, copper, platinum, and so forth. Perfect contact is assumed between the film and the coaxial conductors.
3. The diameter values of the coaxial lines are very small compared with a wave length.
4. The coaxial line terminating impedance (that is, the receiver input impedance) is of such a value that the impedances along the line, beyond the film, remain fairly high, that is, 1 ohm or higher.

The general field-theory approach[2,3] treats the solid metallic disk (or annulus) as a section of a coaxial line with an intrinsic impedance corresponding to that of the particular metallic medium. Under the preceding assumptions the fields in all the coaxial line sections are that of the dominant transverse electromagnetic (TEM) plane wave mode of propagation.

Let ϵ_m, μ_m, σ_m and η_m designate respectively the permittivity, permeability, conductivity, and intrinsic impedance of an infinite metallic medium. Under the preceding assumptions

$\omega\epsilon_m << \sigma_m$

where

$\omega = 2\pi f$

For copper at 1,000 megacycles

$\omega\epsilon_m = 0.17$ mhos per meter

as against

$\sigma_m = 5.7 \times 10^7$ mhos per meter

Therefore[3]

$$\eta_m \cong (1+j)\sqrt{\frac{\omega\mu_m}{2\sigma_m}} \tag{1}$$

The propagation constant in the same medium is

$$\gamma_m \cong (1+j)\sqrt{\frac{\omega\mu_m\sigma_m}{2}} \tag{2}$$

The characteristic impedance of a coaxial line with a solid conductor as a propagation medium is

$$Z_0 = \eta_m \frac{1}{2\pi} \ln\frac{r_1}{r_2} \cong (1+j)\frac{(\omega\mu_m)^{1/2}}{2\sigma_m}\,\frac{1}{2\pi}\ln\frac{r_1}{r_2} \tag{3}$$

where r_1 and r_2 are respectively the large and small radii of the conductors.

The d-c resistance of the annulus of thickness d is

$$R_m = \frac{1}{\sigma_m 2\pi d}\ln\frac{r_1}{r_2} \tag{4}$$

and the depth of penetration is

$$\delta = \left(\frac{1}{\sigma_m \pi f \mu_m}\right)^{1/2} \tag{5}$$

Therefore

$$Z_0 \cong (1+j)R_m\frac{d}{\delta} \tag{6}$$

and

$$\gamma_m = (1+j)\frac{1}{\delta} \tag{7}$$

Using well-known transmision line theory, the output voltage V of a line section equivalent to the annulus terminated by an impedance Z_r is given in terms of the input current I_1 by

$$V_r = \frac{I_1}{\dfrac{1}{Z_r}\cosh\gamma d + \dfrac{1}{Z_0}\sinh\gamma d} \tag{8}$$

and the transfer impedance is

$$Z_m = \frac{V_r}{I_1} = \frac{1}{\dfrac{1}{Z_r}\cosh\left[(1+j)\dfrac{d}{\delta}\right] + \dfrac{1}{R_m}\,\dfrac{\sinh[(1+j)d/\delta]}{(1+j)d/\delta}} \tag{9}$$

In this application $Z_r >> R_m$, therefore for all practical purposes

$$Z_m = R_m(1+j)(d/\delta)\,\mathrm{csch}\,[(1+j)d/\delta] \tag{10}$$

or taking the first three terms of the expansion[4]

$$|Z_m| = R_m\left|1 - j\frac{d^2}{3\delta^2} - \frac{7}{90}\frac{d^4}{\delta^4} + \ldots\right| \tag{11}$$

The absolute value of the transfer impedance is therefore equal to the d-c resistance of the annulus to better than

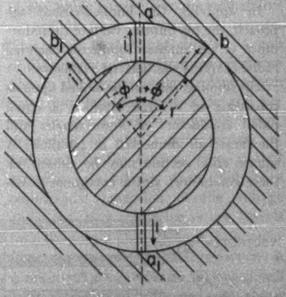


Figure 3. Radial conductor structure approaching that of a solid annular ring

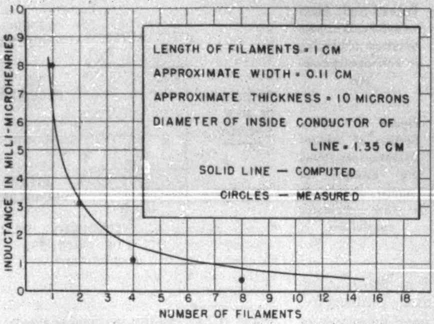


Figure 4 (left). Inductance of a system of tin (Sn) radial filaments short-circuiting the output end of a coaxial transmission line

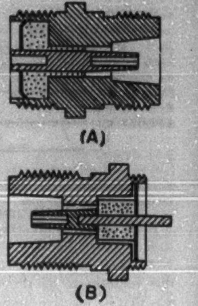


Figure 6 (right.) Cross sections of interchangeable Micropotentiometer resistance element assemblies. Solid black line indicates the metallic film, part of which is the annular resistor

A. Reversible unit using ceramic insulator. (When used with resistance film on the outside of the Micropotentiometer housing, maximum accuracy is assured. When reversed, type N connector is applicable)

B. Unit using glass bushing as insulator

1 per cent for $d/\delta \leq 0.5$, and

$$|V_r| = I_1 |Z_m| \cong I_1 R_m \qquad (12)$$

That this is true can also be shown in a simpler but less rigorous manner, as follows.

The input impedance Z_m' of the line section (Figure 1) having a metal as a dielectric and terminated in an open circuit or in an impedance $Z_r >> Z_0$ is

$$Z_m' \cong \frac{Z_0}{\operatorname{tgh} \gamma_m d} = \frac{R_m (1+j)^{d/\delta}}{\operatorname{tgh} [(1+j)^{d/\delta}]} \qquad (13)$$

For sufficiently small values of d/δ therefore

$$Z_m' \cong R_m \qquad (14)$$

Since for small values of d/δ the current in the annulus is uniform over its thickness to better than 1 per cent, the annulus can be treated as a resistance element is ordinarily treated at direct current, and the voltage drop across it is simply the product of its d-c resistance and I_1.

As was already mentioned, the success or failure of this device depended on the requirement of negligible reactance in the annulus. Therefore all possible substantiating evidence was investigated both analytically and experimentally.

In applying the lumped-circuit approach, the top limiting value of the inductance of an annular conductor carrying radial currents was derived as follows.

The annular ring can be looked upon as consisting of a large number of equal radial conductors of finite length and thickness, all connected in parallel. A square cross section may be chosen for these conductors and a maximum possible value of the ring inductance evaluated by computing the effective inductance of the system of radial conductors. Figure 3 shows four of these radial conductors starting at the periphery of the center conductor and ending at the outer conductor of a coaxial line. The coaxial conductors are at right angles to the radial elements and therefore have no mutual inductance with them. The return path for the current in the annulus is assumed far enough away to avoid any influence on the radial elements. Complete penetration of the current, that is, uniform current density in these elements is assumed at all frequencies. Consequently all deductions based on low-frequency analysis may be used at the higher frequencies at which current density remains essentially uniform.

Each radial element has self-inductance and mutual inductance with all the other elements. Assuming an even number of elements, for each element b on one side of a given element a there is another element b_1 located at an angular symmetry which will cancel the mutual effect of b with a. Thus the effective inductance of a will be

$$L_a = L_a' + M_{ab} + M_{ac} + \ldots + M_{ab_1} + M_{ac_1} + \ldots + M_{aa_1} = L_a' + M_{aa_1} \qquad (15)$$

where L_a^1 is the self-inductance of element a and M_{aa_1} is the mutual inductance between element a and a_1 located diametrically opposite a.

Let d be the thickness of the element, and r be the radius of the inner coaxial conductor. Then the total inductance of the annulus, L_A, will be equal to the inductance of all the elements in parallel, or it may be considerably less than this value because of the additional parallel elements not accounted for, which complete the solid annular ring. Thus

$$L_A \leqq \frac{d}{2\pi r}(L_a' + M_{aa_1}) \qquad (16)$$

The self-inductance of a square bar is[5]

$$L_a' = 0.002l \left[\log \frac{2l}{0.447d} - 1 + \frac{0.447d}{l} \right] \text{ microhenrys} \qquad (17)$$

where d = thickness of the square bar in centimeters and l = its length in centimeters.

The mutual inductance between elements a and a_1 will have a negative sign, because they are connected in series bucking, and is given by[6]

$$M_{aa_1} = -0.002[(2l+2r)\log(2l+2r) + 2r \times \log 2r - 2(l+2r)\log(l+2r)] \text{ microhenrys} \qquad (18)$$

Computations for a typical annular ring for resistances of the order of one milliohm, having

$l = 0.1$ centimeter
$d = 2.5 \times 10^{-4}$ centimeters
$r = 0.5$ centimeter, result in
$L_a \cong 0.45$ millimicrohenry
$M_{aa_1} \cong 8$ micromicrohenrys
$L_A \leqq 0.035$ micromicrohenry

The series reactance of this inductance is about 22 microhms at 100 megacycles which seems entirely negligible even at considerably higher frequencies, it being in quadrature with the resistive component. The actual series inductive reactance may be much lower, as already indicated, and seems still less important as the annular resistance is increased. The validity of the lumped circuit approach was verified experimentally. Equations similar to 16, 17, and 18 were used to compute the inductances of

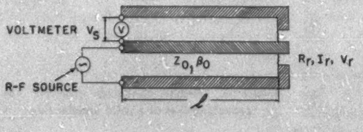


Figure 5. Schematic representation of "coaxial" type Micropotentiometer

symmetrical systems of radial tin (Sn) foil filaments of a centimeter in length and approximate width and thickness of 0.11 centimeter and 10 microns respectively, shorting the output end of a slotted measuring transmission line. The inductances were measured at 200 and 400 megacycles. Figure 4 shows a fairly good agreement between computed and measured values of inductances of these conductors, with a tendency of the measured values to diminish somewhat faster than the computed, with increasing number of filaments.

The only conclusive experimental evidence verifying the preceding analytical predictions was obtained by comparing voltage drops across annular elements with those of a voltage standardization bolometer bridge.[7] These results will be discussed later. Some further supporting evidence was found in the apparent failure to detect inductance in shorting disks of precision slotted transmission lines. None seemed to have been reported in the past anywhere in the field of slotted lines, resonance lines, nor in coaxial resonant cavities. However the impedance values here under consideration require a precision of measurement of voltage-node displacements of the order of a micron or better; such a precision was hardly available to anyone to date.

RANGES OF VOLTAGE AND FREQUENCY

The general conclusion that can be drawn from the preceding is that annular resistance elements of the type described can be used without frequency corrections over the entire ultrahigh-frequency range and higher, and indeed up to the frequency for which the metal-film thickness required is insufficient to maintain a homogeneous, continous metallic medium. This limit (about 0.05 of a micron) indicates that the top frequency of the Micropotentiometer application is dictated by the current indicators and higher modes of propagation rather than by the resistance elements.

At any frequency the resistance of the annulus will remain the same as at direct current to within 1 per cent or better as long as the thickness of the conductive film does not exceed one-half the depth of penetration at that frequency for a thick conductor of the same material. If the film thickness is equal to the full depth of penetration, there may be an error of about 3 per cent in the r-f output voltage. It is necessary to select materials of proper conductivity and choose annular diameters of proper ratios and practical values to obtain various resistances desired to a given top frequency. One might call attention to the fact that the resistance is a function of diameter ratios and not of diameter values. Thus annular resistances of a fraction of a milliohm to 1 ohm and even higher may be used over very wide frequency ranges.

The voltage range of a given resistance element is clearly a function of the power it can dissipate in a particular physical arrangement. However, the power levels in question are generally of such a low level and the annular films are in such close proximity to large and efficient heat conductors that the voltage range of a Micropotentiometer with a given resistance element seems limited at the present time by the current monitoring elements rather than by power dissipation of the resistance films. No appreciable resistance changes were observed on 1- to 10-milliohm silver elements for currents up to an ampere.

Any means to indicate current accurately and independently of frequency are usable with the Micropotentiometers provided they can be physically located in such a manner that the same current is passing through the annular resistor as through the current monitor. This condition becomes more difficult to fulfill as frequency is increased. However, the fact that the annular resistances are very low renders the application of conventional r-f current indicators considerably less critical than in their application with high resistance elements. Thus thermoelements and thermistors were successfully used, as will be shown later, and other bolometer type and thermoelectric elements are applicable for frequencies approaching 1,000 megacycles and perhaps higher.

The low limit of the voltage range is controlled by the smallest practical diameter ratios of the annular elements, by the conductivity of the materials, by the accuracy with which low d-c resistance values can be determined as well as by the accuracy with which low current values can be measured. For frequencies up to the very-high-frequency range the low limit to date seemed of the order of 0.1 of a microvolt; for higher frequencies it was of the order of 1 microvolt.

GENERAL DESIGN FEATURES

In designing assemblies of components and housing of Micropotentiometers only one somewhat critical requirement has to be fulfilled. The resistance films comprising the annular elements and having various thicknesses down to a micron and lower have to be in perfect continuous contact with the outside and inside conductors and must remain mechanically rugged and stable. Other requirements are of the ordinary variety which can be readily met. One of these requirements concerns the proximity of the current monitoring element to the annulus and is not critical because there is a current loop at the annulus; the current falls off as a cosine function with distance towards the current monitor (or r-f power source). This feature tends to equalize the current distribution along heaters of thermoelements as well, and thus extends the frequency range. For example, a distance of 2 centimeters from the resistive film to the current indicator may cause a 1-per-cent error at 300 megacycles. With thermistors as current monitors this question seems of no consequence, even over the entire ultrahigh-frequency range since thermistor beads having diameters of 15 mils can be placed right at the annulus.

Another noncritical requirement concerns the length of the output coaxial connector of the Micropotentiometers. This connector may be considered an integral part of the feeder supplying the standardizing voltage to a receiver or any other monitor. For best results the annular resistance film should be placed in the output plane of the Micropotentiometer and the connectors to the monitor and to the equipment being calibrated in terms of the Micropotentiometers should be mechanically matched. In case this cannot be done, then the annular resistance film should be placed as close as possible to the output plane. The accuracy will then depend on the length of the Micropotentiometer connectors and the input impedance to the monitoring feeder. To be more specific, the output of the Micropotentiometer may be affected in two ways. The first is the actual increase of the shunt admittance across the annulus; since the resistance of the latter is very low the probability of trouble from this cause is remote. The second is the transformation action of the short connector from the resistance film to the physical output plane which may contribute an error of several per cent in the ultrahigh-frequency range, depending on the monitor-feeder impedance. In the extreme case, if that impedance is infinite, there will be a cosinusoidal reduction of the voltage along the short connector as one travels from the output plane towards the annular film; for example, at 300 megacycles a connector 1.5 centimeters in length will have a voltage about 1 per cent higher at the output plane as compared with

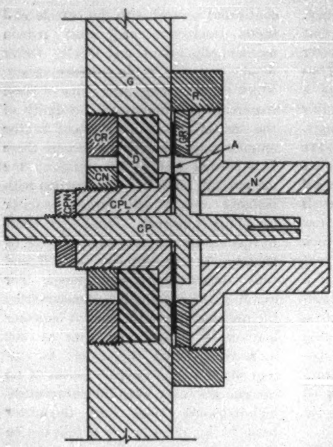


Figure 7 (left). Cross section of interchangeable Micropotentiometer resistance element assembly. Solid black line indicates resistance ring clamped between electrodes

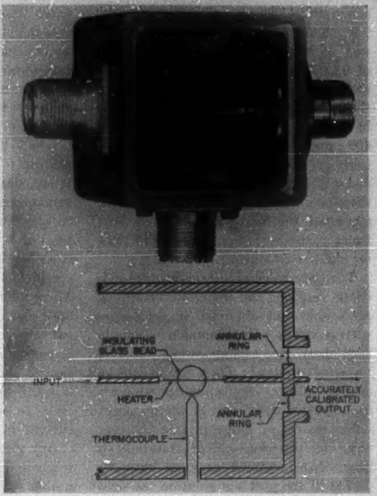


Figure 9 (right). Micropotentiometer employing a thermoelement as a current indicator

the standardizing voltage across the annulus. As the monitoring-feeder input impedance decreases, this error will be reduced and will be zero when this impedance is equal to the characteristic impedance of the connector.

One might point out here a practical advantage of these devices for universal standardization purposes. A known resistor of reliable r-f characteristics may be used as the coaxial center conductor of the output terminal. This would form a well-defined readily reproducible impedance of the source of standardizing r-f voltages when reliable voltage generators with given output impedances are required. Conventional standard "dummy antennas" may of course be used instead, with full certainty that the source impedance in this case is that of the dummy antenna only.

Types of Micropotentiometers

The basic principle of the Micropotentiometers lends itself to balanced circuit applications as well as to unbalanced. Since development work, so far, has been limited to unbalanced requirements, only the latter are discussed here.

The only other classification of these devices as to general type, aside from the preceding, is based on current measuring methods. These affect the basic physical structure of the units. All units using thermoelements may be essentially of the same construction and may have interchangeable annular and current indicating elements to cover a wider voltage range. Insulated-type thermoelements preferably should be used. In this type the structure requires only the thermocouple output terminals in addition to r-f input and output terminals. On the other hand, types employing bolometers (for example thermistors) or directly heated thermocouples require either r-f chokes or d-c blocking condensers, and in some cases, both.

One type, radically departing from the preceding and having some desirable features, is a Micropotentiometer using a reliable voltmeter instead of a current indicator. Figure 5 shows this modification schematically. R-f power is fed into a section of an air or solid dielectric coaxial transmission line having negligible losses. A calibrated vacuum-tube voltmeter connected at the input to the line measures the input voltage. The output end of the line is terminated in a solid metallic disk containing an annulus of one-half the penetration thickness at the highest frequency of interest. Then, to a very good approximation

$$V_s = I_r Z_0 \sin \beta_0 l \tag{19}$$

and

$$V_r = I_r R_r = V_s \frac{R_r}{Z_0} \csc \beta_0 l$$

$$= V_s K_1 \csc K_2 f \tag{20}$$

where

R = the resistance of the annulus, ohms
V_r and V_s = the output and input voltages respectively, volts
I_r = the current in the short circuit, amperes
Z_0 = the characteristic impedance of the line, ohms
λ_0 = the wave length in the line, meters
$\beta_0 = 2\pi/\lambda_0$ = phase constant of the line
l = length of the line in meters
f = frequency in cycles per second

Since K_1 and K_2 are both constants

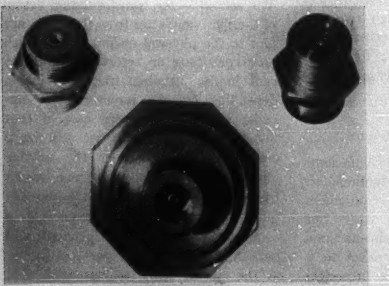


Figure 8. Micropotentiometer interchangeable resistance assemblies

A, B. Units using glass or ceramic bushings and evaporated and plated or fired-on films

C. Unit using clamped carbon disk

for a given air or solid dielectric line, one can compute and plot V_r versus frequency for $V_s=1$ volt to facilitate the determination of V_r over wide frequency and voltage ranges. A further simplification is obtained by using check frequencies for which the line length is an odd multiple of $\lambda_0/4$, in which case

$$V_r = \frac{R_r}{Z_0} V_s \tag{21}$$

It can be shown that equation 21 holds for any value of R_r, thus a wide range of voltages is available at the output of this type of Micropotentiometer; for example, with $Z_0=50$ ohms, $V_s=0.1$ volt and interchangeable R_r elements of 0.5 to 100 milliohms, one can obtain V_r values of 1 to 200 microvolts.

Practical Micropotentiometers and Verification of Performance

The ideal Micropotentiometer should have a construction whereby the annular resistance element constitutes an integral part of the inner or outer surface of the relatively heavy metallic enclosure. There are numerous ways to approach this ideal and many of them were tried with various degrees of success. Details of construction and fabrication processes are beyond the scope of this discussion. However, three different satisfactory methods will be indicated on the basis of up-to-date experience. Low resistance elements, of the order of 0.2 to several hundred milliohms, were constructed in one of two ways. First, ceramic or glass cylinders were sealed as insulators between the two concentric conductors. For this sealing, metallic paints or pastes (silver, platinum, gold) were fired on the cylindrical surfaces of these insulators; the actual sealing was accomplished by soldering. An annular resistance element was then formed at one end of the coaxial assembly by evaporating and plating a given metal over its entire surface. Second, the same process was used as in the first case except that the metallic paints or pastes were also fired on one of the ends of the glass or ceramic insulators to form the annular resistor; evaporation and plating were thereby eliminated. A third method was applied successfully only to elements of the order of 1 ohms in resistance; carbon (deposited on bakelite) disks were clamped in a solid coaxial assembly; silver painted rings were used as conducting electrodes for the disk at the clamped surfaces. One modification of the low resistance elements made use of commercial Kovar-glass sealed terminals; metallic films fired on, or evaporated and plated over one side of the terminals formed the resistance elements; the terminals were then soldered into an appropriate housing. Figures 6, 7, and 8 show cross-sectional and photographic views of these elements and Figure 9 shows an assembled unit employing a thermoelement.

Once a Micropotentiometer is assembled, all that is necessary in order to determine its voltage output is to measure the d-c resistance of the annular film and to calibrate the current indicating element on direct current. However, in the up-to-date stages of this development, it was necessary to verify the results in terms of other independent methods at all frequencies. Reliability of new units can, of course, in the future be checked against older units with little difficulty. Agreement tests were conducted against the voltage standardizing bolometer bridge, mentioned previously[7] and precision wave guide below cutoff attenuators. These tests indicated agreement well within over-all experimental errors, that is ±1 per cent to about 50 megacycles and ±3 per cent to 300 megacycles. Measurements at higher frequencies have been conducted so far only on the clamped 1-ohm units and resulted in agreements of ±5 per cent to 900 megacycles.

Application of Micropotentiometers

Though the primary objective of these Micropotentiometers is to eradicate the wide uncertainty in the absolute values of voltages in the microvolt range, they can be used for numerous other purposes. The more obvious applications may be briefly indicated as follows:

1. Reference standard for accurate voltages in the range previously indicated for calibration of voltmeters, signal generators, field intensity meters, and so forth. The voltage and frequency range seems to be limited primarily by current-indicating facilities.

2. Sources of accurate voltages for direct use in place of conventional signal generators. The user has complete freedom to vary the internal impedance of these sources for various requirements with a high degree of certainty in the values of this impedance. A compact assembly incorporating an oscillator and Micropotentiometer may be constructed as a standard voltage (signal) generator.

3. Calibration of attenuators directly in terms of voltage ratios.

4. Determination of performance of current indicators, such as of various types of thermoelements.

5. Determination of output impedances of r-f sources. Resistance elements of the Micropotentiometers may be used to short-circuit the output of a source; the voltage output of these elements is then a measure of the short-circuited current of the source and consequently of its internal impedance.

6. Calibration of modulation meters. This application is common to all devices having an accurately known correlation between output and percent modulation, for example, the output of the couple of the thermoelement. Attention may be called here to the fact that presence of harmonics in the r-f carrier will normally have a negligible effect on the couple output, whereas the effect may be very appreciable on a vacuum-tube voltmeter usually employed with conventional voltage generators.

7. Applications where voltages in the microvolt range are desired without the usual Johnson noise of higher impedance sources present.

8. Applications where constant voltage sources are required, for example for Q-measurement circuitry.

9. Applications where a single device to cover the entire frequency range from zero to 300 megacycles and higher is essential.

References

1. RADIO-FREQUENCY MICROPOTENTIOMETERS. *Technical News Bulletin*, National Bureau of Standards (Washington, D. C.), volume 35, March 1951, pages 33–34.

2. ELECTROMAGNETIC WAVES (book), S. A. Schelkunoff. D. Van Nostrand Company, Inc., New York, N. Y., 1943, page 280.

3. NETWORKS, LINES AND FIELDS (book), J. D. Ryder. Prentice Hall, Inc., New York, N. Y., chapters 8 and 10, 1949.

4. TABLES OF COMPLEX HYPERBOLIC AND CIRCULAR FUNCTIONS, A. E. Kennelly. Harvard University Press (Cambridge, Mass.), second edition, 1927, page 231.

5. INDUCTANCE OF LINEAR INDUCTORS, E. B. Rosa. *Bulletin*, National Bureau of Standards (Washington, D. C.), volume 4, number 2, 1907–08, page 313.

6. *Circular C74*, National Bureau of Standards (Washington, D. C.), 1937, page 270.

7. A BOLOMETER BRIDGE FOR STANDARDIZING RADIO-FREQUENCY VOLTMETERS, M. C. Selby, L. F. Behrent. *RP 2055*, *Journal of Research*, National Bureau of Standards (Washington, D. C.), volume 44, 1950, page 15.

◆

Discussion

Donald M. Hill (Boonton Radio Corporation, Boonton, N. J.): The micropotentiometer appears to be a worth-while contribution. Some time after the micropotentiometer was announced we had an application for it as the coupling impedance in a Q meter. After conferring with Mr. Selby concerning the methods of construction, we constructed a number of the resistors by firing a platinum-gold alloy on ceramic discs. These were tested over the frequency range from zero to 50 megacycles and found to be suitably free of inductance.

The temperature coefficient of the combination of an annular resistor and a thermocouple was found to be 0.11 per cent per degree centigrade.

One point which has perhaps not been sufficiently emphasized by Mr. Selby is that in order for the inductance to be negligibly low the resistive film must have complete axial symmetry. Until a manufacturing method can be devised which insures perfect symmetry, it will be necessary to test each resistor individually.

Myron C. Selby: The author wishes to thank Dr. Hill for his encouraging discussion. One can hardly question Dr. Hill's emphasis of the axial symmetry of the resistive film because it is a basic requirement of this development. Unfortunately we at the National Bureau of Standards have had no opportunity as yet to correlate degree of departure from symmetry with accuracy. Annular elements of superior symmetry were fabricated by means of evaporation and plating. However, there was no evidence to date that this fabrication resulted in units of higher average accuracy compared with others.

Load-Dropping Tests on a Large Ignitron Rectifier Installation

S. J. POPE
MEMBER AIEE

J. K. DILLARD
ASSOCIATE MEMBER AIEE

C. R. MARCUM
MEMBER AIEE

A SERIES of load-dropping tests on ignitron rectifier circuits was conducted during March 1952 on an aluminum potline at the Mead (Wash.) Works of the Kaiser Aluminum and Chemical Corporation. The purpose of these tests was to obtain fundamental data on current and voltage surges encountered in a large ignitron rectifier installation when the potline load was tripped off by each of the following methods.

1. By a-c circuit-breaker operation.
   a. With auxiliary power maintained on the rectifier.
   b. With simultaneous tripping of auxiliary power.
   c. With simultaneous short-circuiting of all ignitor-firing circuits.
2. By automatic phase back to reduce the potline voltage to approach polarization value, followed by tripping of cathode circuit breakers.
3. By tripping all cathode breakers simultaneously.

Such fundamental data were desired in order to determine the best operating procedure to be used in dropping potline load in proposed higher capacity ignitron installations. In addition, it was desired to establish whether or not measures should be taken to protect existing installations against equipment failure from switching surges.

History

An outstanding development of World War II was the enormous increase in the use of rectifiers for the electrolytic reduction of light metals, both in this country and in Canada. During the war, a number of rectifier potlines operating at approximately 60,000 amperes and 650 volts were placed in service.[1] To meet the needs of the nation's defense, the use of ignitron rectifiers is again rapidly expanding both in total installed capacity and the relatively large number of units operating in parallel. Since the output of metal from the reduction process is almost directly proportional to the current through the pots, potlines which will carry 72,000 amperes at 750 volts are under construction, and lines are being proposed which will operate in excess of 100,000 amperes and 750 volts.

On rectifier installations placed in service during the war, potline loads are usually dropped by master trip of the cathode circuit breakers.[2] Some installations were originally set up to drop load by short-circuiting the ignitors to stop firing, but this method was abandoned because arc backs frequently accompanied the switching operation. On earlier installations, it was observed that dips of the order of 25 per cent in supply voltage caused the firing of the ignitrons to be erratic, resulting in unbalanced division of load between parallel rectifiers. Because of this phenomenon, little consideration was given to dropping load by opening the a-c circuit breaker. An advantage in master-tripping the cathode breakers was that auxiliary power supplying the firing circuits was invariably taken from the load side of the a-c circuit breaker, and it seemed to be desirable to keep the firing circuits energized when load was dropped. These factors led to the adoption of this method in spite of the stress placed on the cathode breakers, the last breaker to clear being required to interrupt approximately one-half of the potline current. In proposed higher capacity installations it would be desirable to relieve cathode breakers of this burden.

There have been a few rectifier transformer failures on installations made early in the war.[3] Some transformer failures were clearly the result of mechanical stresses, probably due to repeated arc backs, as evidenced by Figure 1 which shows mechanical distortion of the winding in the region of the failure. Magnetic forces resulting from short circuits or arc backs may have been responsible for dislodging the pressboard spacers between windings. Once the spacers between coils became dislodged, the high-voltage and low-voltage coils could shift with respect to each other, resulting in greatly increased magnetic forces which cause a part of the winding to collapse and burn out. As a result, replacement windings were constructed with additional bracing to withstand expected mechanical stresses resulting from arc-back currents, and subsequent windings were designed to withstand substantially higher mechanical forces.

Some failures, such as the one shown in Figure 2, do not show any mechanical distortion. This led to the consideration that these coils may have failed because of breakdown of the insulation by transient surges, although no other evidence of the existence of surges in such installations has ever been found. A possible source of excessive voltage surges is the rapid opening of circuit breakers either in normal switching operations or to clear faults. Another possibility is the phenomenon of arc starvation—the sudden interruption of current due to a deficiency of ions in the arc path. While it

---

Paper 53-65, recommended by the AIEE Electronic Power Converters Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted June 16, 1952; made available for printing December 4, 1952.

S. J. Pope is with the Kaiser Aluminum and Chemical Corporation, Spokane, Wash. J. K. Dillard and C. R. Marcum are with the Westinghouse Electric Corporation, East Pittsburgh, Pa.

The authors gratefully acknowledge the contributions of U. B. Graff of the Kaiser Aluminum and Chemical Corporation and of the engineers of the Westinghouse Electric Corporation who participated in the tests and helped prepare the final report.

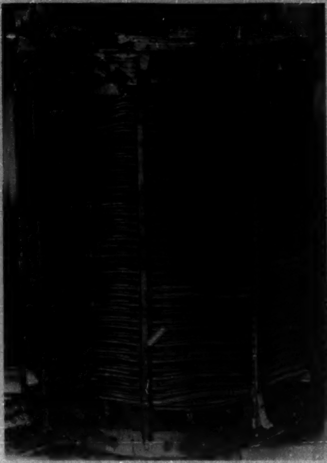


Figure 1 (left). Rectifier transformer failure showing mechanical distortion

Figure 2 (right). Rectifier transformer failure with no mechanical distortion

is doubtful that arc starvation occurs on ignitrons, a similar occurrence, known as arc extinction, has been observed.[4] Arc extinction in ignitrons is the tendency for the tubes to stop conducting abruptly when small currents, of the order of a few amperes, are flowing. Because of the circuit inductance, the sudden interruption of potline current might produce an extremely high voltage of very short duration.

Transformers on later installations were designed to withstand substantially higher mechanical forces and with higher impulse levels than was the case on the first installations. There have been no failures on these transformers.

Although little attention has been directed to methods of dropping load other than by master-tripping of cathode circuit breakers, a number of investigations have been made, with limited success, in an attempt to run down the cause of transformer failures. For these reasons field tests were undertaken. Tests were expected to lead to the most desirable method of dropping potline load. In addition, if excessive surges do result from switching operations, these could be detected. Such knowledge would be useful to the industry in properly co-ordinating apparatus used in electrochemical installations and in the operation of rectifier circuits and equipment.

Discussion of Tests

Several factors made the Mead Works desirable for the performance of these tests. In the first place, with the method of phase shifting employed at Mead, it was possible to reduce potline voltage by automatic phase control just before tripping the cathode circuit breakers. Secondly, a transformer failure had occurred at Mead, and it was felt that if circuit conditions were the cause of failures these might be discovered. In addition, connections to the neutrals of the four secondary windings are brought out from the transformer tank, permitting voltage measurement between each terminal of the transformer winding and the neutral of the opposite wye. Since these two points are physically adjacent on each phase of the secondary windings, the insulation separating them is the most vulnerable point to surges entering the transformer.

A 1-line diagram of the rectifier potline used in making the field tests is shown in Figure 3. The potline consists of 12 ignitron-rectifier assemblies supplied from six rectifier transformers with double primaries and quadruple-wye-connected secondaries. Figure 4 is a schematic diagram showing one transformer supplying two rectifier assemblies. The circuit has been simplified by representing two parallel ignitron tubes and the associated balance coil by a single tube. The rectifier ignitrons are fired by a saturating reactor-type firing circuit with a phase-shifting network for adjusting the output voltage.[5] The potline is operated as a 36-phase unit by connecting the primaries of three rectifier transformers in delta, three rectifier transformer primaries in wye, and then supplying four of the units through phase-shifting transformers. Switchgear consists of high-speed anode circuit breakers, semihigh-speed cathode circuit breakers, and a plain-break a-c oil circuit breaker of the type used on early rectifier installations. Normal operating

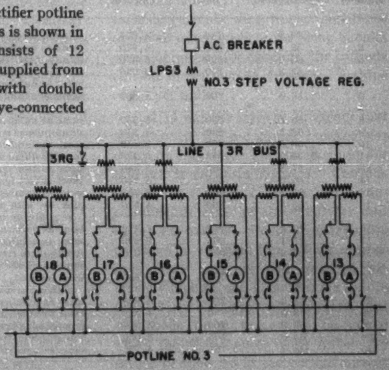


Figure 3. One-line diagram of test potline

current of the potline is about 60,000 amperes at 700 volts.

The tests are listed in chronological order in Table I. In this table, meter readings taken from the operator's board just prior to the test are recorded. In the case of tests 15 and 16, in which potline load was picked up, the readings refer to conditions immediately following the test.

LOAD INTERRUPTION BY A-C CIRCUIT-BREAKER OPERATION WITH AUXILIARY POWER MAINTAINED

Tests 1, 2, 6, and 8 were made in which potline load was dropped by tripping the a-c circuit breaker while maintaining auxiliary power in order to keep the firing circuit energized while dropping the load. In each test involving a-c circuit-breaker operation, simultaneous oscillograms were made of the following: two line-to-line voltages on the transformer primary; six cathode currents, assembly *A* from each of the six rectifier units connected to the potline; all six secondary phase currents of rectifier 13-A; potline current; and potline voltage. Maximum crest values of currents taken from the oscillograms are tabulated in per unit of normal value under load in Table II. The oscillogram for test 8 is shown in Figure 5.

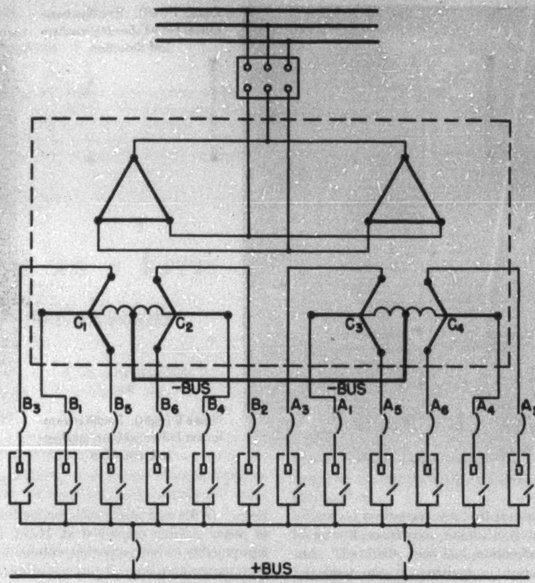


Figure 4. Schematic diagram of transformer connections

In one case, test 1, two restrikes occurred in the a-c circuit breaker, resulting in two oscillations in potline voltage of maximum value of 1.0 per unit before final interruption. Corresponding oscillations occurred in the recorded currents, but the maximum cathode overcurrent was 1.3 per unit even in this adverse case. No secondary phase current exceeded 1.0 per unit.

Tripping the alternating voltage, as was done in these four tests, removed any tendency for currents between parallel rectifiers to become unbalanced. In addition, the current in the transformer secondary windings was confined to two windings 180 degrees out of phase. For example, in Figure 5, the first phase left holding current, following a-c breaker trip, was phase 6, phase 1 current was reduced slightly, and phase 2 did not conduct. Because of the inductance of the circuit, the potline voltage was forced to a slightly negative value as the breaker attempted to interrupt the flow of current in phases 6 and 1, leaving a positive voltage on phase 3 which forced it to conduct. After the current in phase 1 commutated to phase 3, the arc was extinguished in the a-c circuit breaker without difficulty since only phases 6 and 3, which are 180 degrees out of phase, were conducting. The potline voltage rose from a slightly negative value to zero, and the cathode current in all units decayed smoothly.

The 6-phase ripple on the potline voltage trace prior to load interruption was due to bus drop, the voltage being measured on the bus where it leaves the rectifier room. As long as two or more ignitrons are conducting simultaneously following interruption of the a-c circuit, the potline is effectively short-circuited except for arc drop in the tubes. This phenomenon accounts for the potline voltage falling almost to zero for several cycles.

Table I. Meter Readings on Operator's Board at Time of Test

Test Number	Kiloamperes	Volts D-C	Cathode Amperes	Remarks
1	64.0	710	4,970	trip a-c breaker, auxiliary power left on
2	56.6	695	4,970	same as number 1
3	68.0	690	5,370	steady-state timing test
4	62.3	710	5,030	same as number 3
5				same as number 3
6	56.6	700	5,030	same as number 1
7				same as number 3
8	56.6	690	5,350	same as number 1
9	62.3	690	5,350	simultaneous trip a-c breaker and auxiliary power
10	56.6	700	5,140	same as number 9
11	58.9	685	5,240	same as number 9
12	57.8	705	5,350	same as number 9
13	60.0	690	5,140	short-circuit ignitors, then trip a-c breaker
14	61.2	690	5,350	same as number 13
15	55.5	700	4,810	pick up with a-c breaker, auxiliary power on; regulator in normal position
16	49.8	710	4,710	same as number 15
17	61.1	695	5,290	same as number 13
18	60.4	695	5,210	steady-state timing test
19				same as number 18
20	62.0	700	5,480	load reduced by phase control before tripping cathode breakers
21	62.0	700	5,260	same as number 20
22	58.0	710	4,950	same as number 20
23	57.0	700	5,000	trip cathode breakers, *13-A* delayed 1 cycle
24	65.0	705	5,060	simultaneous trip of all cathode breakers
25	53.4	705	4,750	same as number 24

Table II. Maximum Crest Values in Per Unit for Tests Involving A-C Circuit-Breaker Operation

Test Number	Cathode Currents						Phase Currents						Potline	
	13	14	15	16	17	18	1	2	3	4	5	6	Volts	Amperes
1	1.31	1.0	1.17	1.33	1.33	1.21	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
2	1.0	1.0	1.0	1.0	1.14	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
3	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
4	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
6	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
7	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
8	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
9, 10*														
11	1.78	1.46	1.14	1.5	1.47	1.61	2.08	1.0	1.23	1.0	1.0	1.29	1.0	1.0
12	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
13	1.36	1.22	1.0	3.5	1.26	1.0	1.39	1.0	2.50	1.0	2.5	1.25	1.0	1.0
14	6.0		1.0	3.6	1.56	1.0	1.85	2.08	7.85	5.16	6.34	1.46	1.22	1.0
15	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.18	1.0
16	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
17	3.43	1.13	1.0	3.8	2.34	1.8	1.48	5.5	4.61	7.23†	5.13	6.82†	1.53†	1.0

* No oscillogram records for tests 9 and 10.

† Maximum crest values occurred during reversed current.

Table III. Maximum Crest Values in Per Unit for Tests Involving Cathode Circuit-Breaker Operation

Test Number	Voltage from Cathode to		Voltage from Negative Bus to		13-A Rectifier		Voltage from Neutral of Opposite Wye to Anode						Potline		Cathode Breaker
	Anode 1	Anode 2	Anode 1	Anode 2	Volts	Amperes	1	2	3	4	5	6	Volts	Amperes	Volts
18	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	
19	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	
20	1.13	1.0	1.78*	2.26		1.17		1.64		2.04		1.79			
21	1.16	1.0	1.76*	2.24*	1.85	5.56	2.0	2.32*	2.18	1.91	2.2	1.46	1.0	1.0	4.1
22	1.10	1.07	1.76	1.85	1.76	5.18	2.06	2.31*	2.09	2.2	2.21	2.15	1.0	1.0	4.0
23	1.39	1.0	1.61	2.11*	1.53	8.25	2.06*	2.33*	1.0	2.55	2.07	1.0	1.0	1.0	5.22
24			1.0	1.0	1.34	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.46
25			1.0	1.0	1.0	1.57	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.46

* Maximum crest value occurred during reversed current.

As soon as conduction stopped, the voltage rose to the back electromotive force of the potline.

The pip in the primary voltage trace about 5 cycles after trip resulted from the flow of exciting current as the potline voltage rose to the counter electromotive force of the potline.

LOAD INTERRUPTION BY TRIPPING A-C CIRCUIT BREAKER AND AUXILIARY POWER SIMULTANEOUSLY

In tests 11 and 12, the a-c circuit breaker and auxiliary power were tripped simultaneously. Such a test is representative of the way the load would be dropped in case of failure of power in the substation supplying the Mead Works. This is also the usual method of operation at Mead.

Figure 6, the oscillogram for test 11, shows that arcing existed in the a-c circuit breaker for about 1 cycle following trip. This test should be compared with test 1 in which quite similar interruption of the a-c circuit occurred. In both cases there was no current disturbance after the a-c circuit was completely interrupted. Referring to Table II, the maximum overshoot of cathode current was about 1.78 per unit, compared with a maximum of 1.33 per unit recorded in test 1. One secondary phase current reached about 2.08 per unit, however, which was considerably more severe than any observed in test 1. In addition, several cathode and phase currents showed a good deal more disturbance. Practically no arcing occurred in the a-c circuit breaker during interruption of the a-c circuit in test 12, resulting in very clean interruption of phase and cathode currents. There were also no restrikes in tests 6 and 8 in which the firing circuits were kept energized while the a-c circuit breaker was tripped. Comparison of test results on this basis indicates that the disturbances of secondary phase and cathode currents in test 11 were the result of arcing in the a-c circuit breaker rather than simultaneous tripping of auxiliary power.

Indications are that this method of dropping load is satisfactory at this installation.

LOAD INTERRUPTION BY SHORT-CIRCUITING IGNITOR CIRCUITS FOLLOWED BY OPERATION OF A-C CIRCUIT BREAKER

Tests 13, 14, and 17 were made by short-circuiting all ignitor circuits simultaneously followed by tripping the a-c circuit breaker several cycles later.

In test 13, the a-c circuit breaker was tripped about $2^1/_2$ cycles after the ignitor circuits were short-circuited. Although no arcing was observed in the circuit breaker, the potline voltage suffered oscillations between the time the ignitors were short-circuited and the time of tripping the a-c circuit breaker. These oscillations were caused by the alternating voltage applied to the tubes which tended to maintain conduction even though the firing circuits were not energized. As shown in Table II, disturbances in some cathode currents were worse than those observed in tests where the firing circuits were energized; for example, the cathode current in rectifier *16-A* reached a value of 4.0 per unit.

Test 14 was the same as 13 except that operation of the a-c circuit breaker was delayed 7 cycles. The oscillogram for this test is shown in Figure 7. Cathode current in rectifier *13-A* reached an instantaneous value of about 6.0 per unit before the cathode circuit breaker finally operated on overload trip. Some secondary phase currents also reached extremely high values, about 8.0 per unit in phase 3, for example. Erratic behaviour of cathode current was evident on several units, the disturbances being sufficiently severe to show up in the potline current

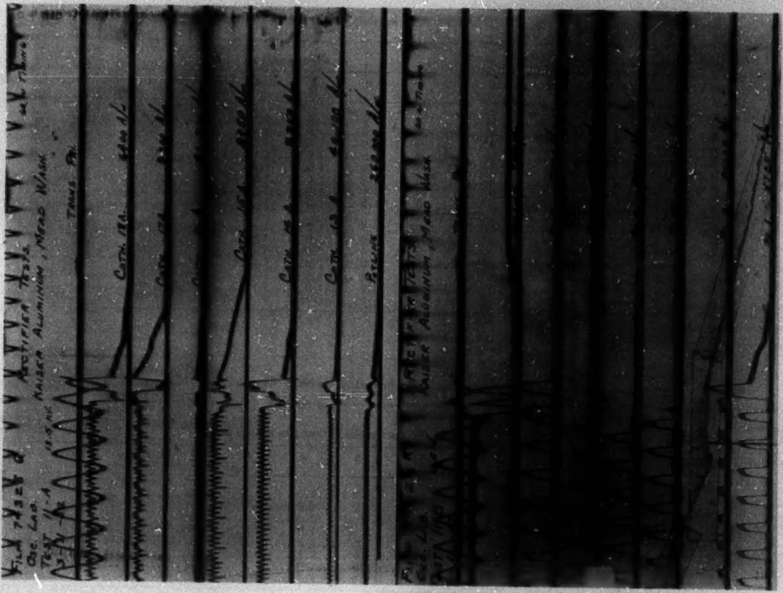


Figure 6. Typical oscillogram dropping potline load by tripping a-c circuit breaker and auxiliary power simultaneously

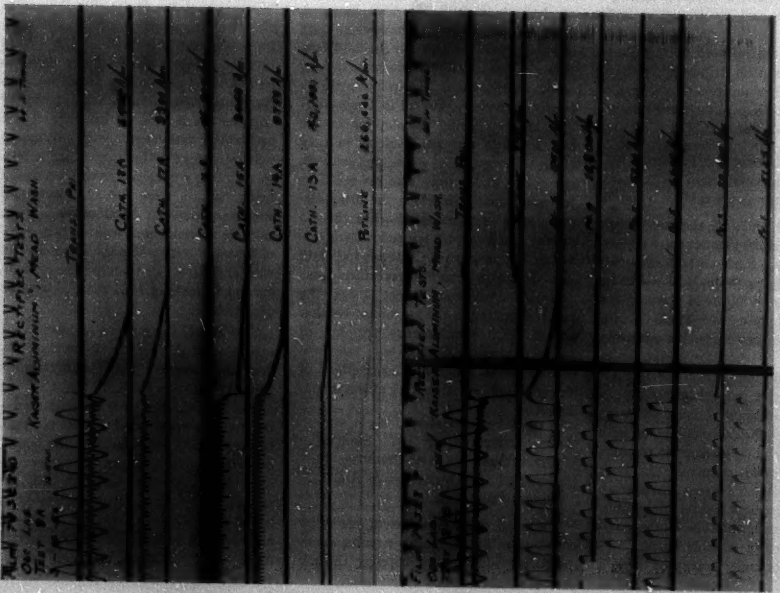


Figure 5. Typical oscillogram dropping potline load by a-c circuit-breaker operation with auxiliary power maintained

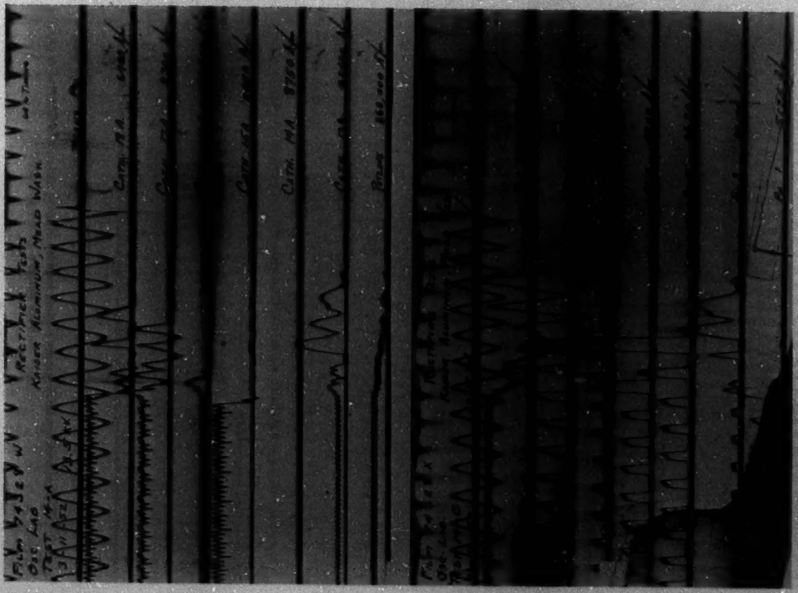


Figure 7. Typical oscillogram dropping potline load by short-circuiting ignitor circuits followed by a-c circuit-breaker operation

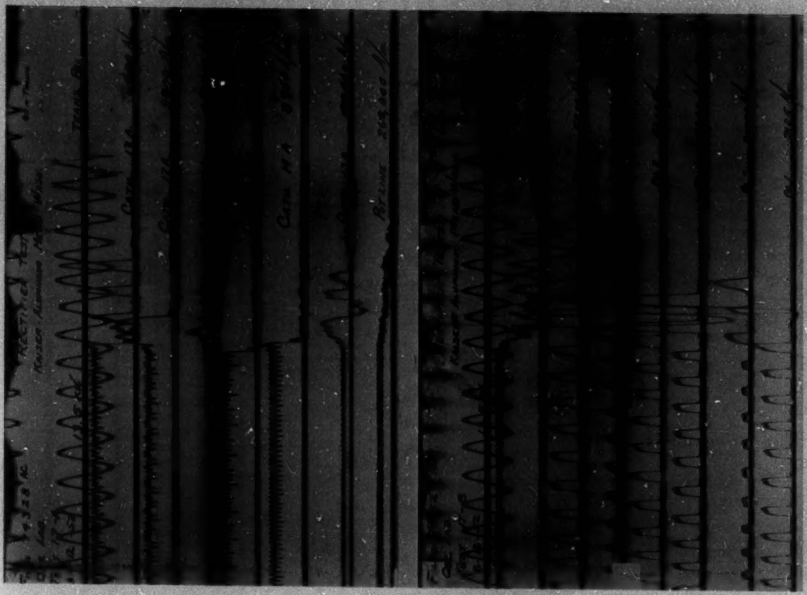


Figure 8. Oscillogram showing natural arc back

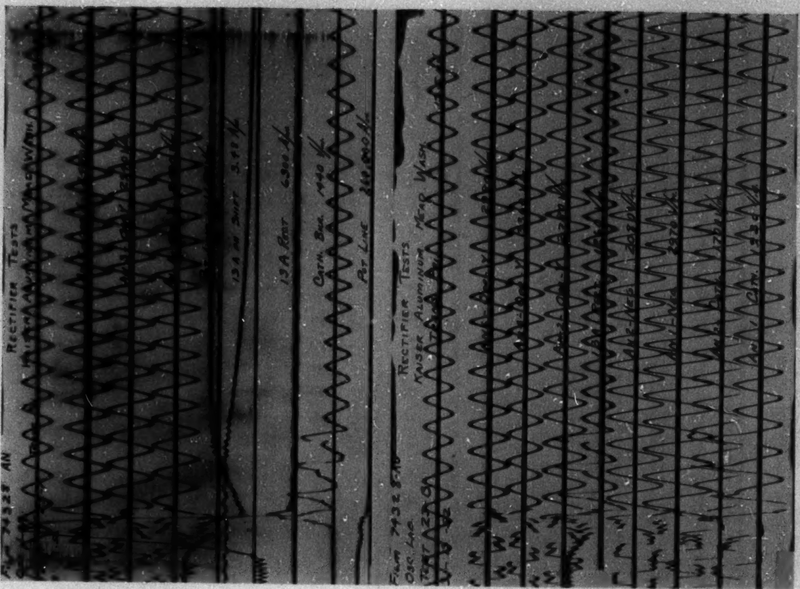


Figure 10. Typical oscillogram dropping potline load by tripping all cathode breakers simultaneously

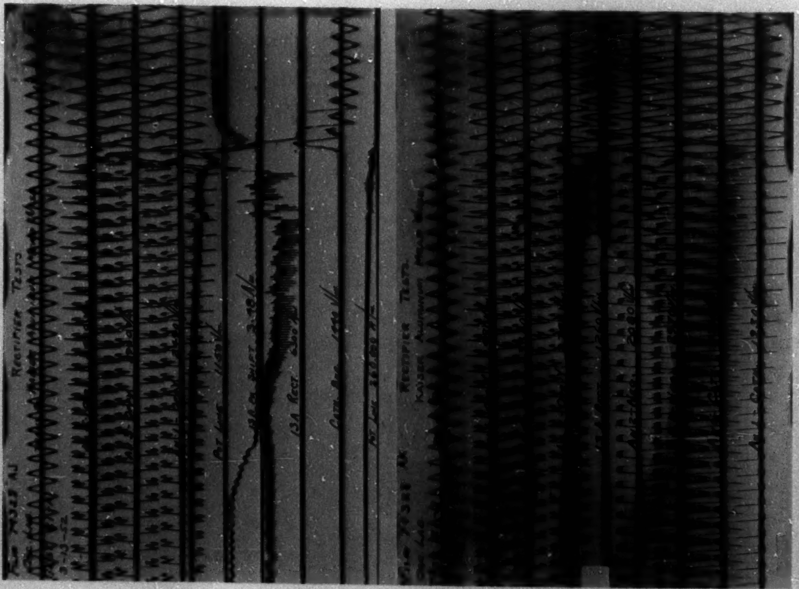


Figure 9. Typical oscillogram dropping potline load by tripping cathode breakers after phasing back ignitors

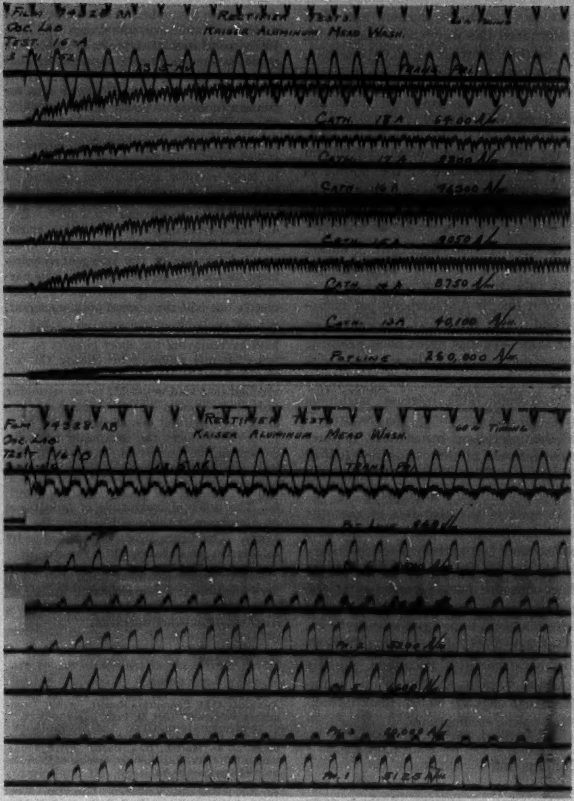


Figure 11. Typical oscillogram picking up potline load by a-c circuit-breaker operation

trace. The 6-phase ripple observed on the potline voltage trace following final interruption of current was picked up across a metering circuit after the cathode breaker in rectifier *13-A* opened.

Test 17 was the same as 14; however, a natural arc back occurred on rectifier *13-A*. Reversal of currents in the cathode circuit and in phases 4 and 6 are clearly shown in Figure 8. This method of load-dropping has been proposed for some rectifier installations. It is obvious that these tests do not support the use of this scheme. In each of these tests, the potline voltage underwent oscillations from the time the ignitors were short-circuited until the a-c breaker opened. Secondary phase, cathode, and potline currents were extremely erratic, arc backs occurring in one test.

LOAD INTERRUPTION BY TRIPPING CATHODE BREAKERS A FEW CYCLES AFTER PHASING BACK THE IGNITORS

In tests 20, 21, and 22 the ignitors were phased back to reduce potline voltage to approach polarization value, followed by tripping of cathode circuit breakers. This method of dropping load is common in Canada and has occasionally been applied in this country.

Operation of the cathode circuit breaker in rectifier *13-A* was delayed approximately 1 cycle in order to force this particular circuit breaker to make the final interruption of potline current. The a-c circuit breaker was kept closed and auxiliary power, energizing the firing circuits, was maintained.

In each test involving cathode circuit breaker operation, simultaneous oscillograms were made of the following: two line-to-line voltages on the primary side of the transformer supplying rectifier *13-A;* voltages between each terminal and the neutral of the opposite wye on the transformer secondary on rectifier *13-A;* two anode-to-negative bus voltages; two anode-to-cathode voltages; voltage and cathode current in rectifier *13-A;* phase-shift current for rectifier *13-A;* voltage across cathode circuit breaker; potline voltage; and potline current. Maximum crest values taken from the oscillograms are tabulated in Table III. A typical oscillogram, the one for test 21, is shown in Figure 9.

None of the oscillograms indicated transient overvoltages on the primary side of the transformer, and no more than 2.0 per unit voltage was evident between any terminal and neutral of opposite wye. The last cathode breaker to clear interrupted approximately one-half of the potline current flowing before interruption. This results because the cathode breakers do not open in perfect synchronization. As the first breaker opens, the current carried by that rectifier is picked up by the other rectifiers because of the inertia effect of the potline inductance. The next breaker must then interrupt a larger share of the instantaneous potline current, and this phenomenon continues until the final breaker out must interrupt the total instantaneous potline current, which by this time has decayed to approximately half value. Cathode current in rectifier *13-A*, in test 21, for example, reached a value of 22,000 amperes, approximately one-half the 52,000 amperes potline current just prior to the test. Interruption of cathode current took place in 0.85 cycle, as evidenced by the duration of arc voltage across the cathode breaker. This is a stringent requirement on the circuit breaker and is the principal objection to this method as there were no serious voltage disturbances.

LOAD INTERRUPTION BY TRIPPING CATHODE CIRCUIT BREAKERS

Tests 23, 24, and 25 were made by tripping all cathode circuit breakers simultaneously without first phasing back the ignitors. Oscillograms were made of the same quantities recorded in the tests where the ignitors were phased back prior to load interruption. The oscillogram for test 23 is shown in Figure 10.

The fact that the potline voltage trace in this test was as smooth as in tests where phasing back was employed indicates that delay of the ignitors does not produce any appreciable improvement in the load-dropping method. The maximum current in the last cathode to clear was larger, 8.0 per unit, but this was expected since phase back reduces the potline current before interruption. Overvoltages were no more severe in this case.

Potline Pickup Tests

Tests 15 and 16 were made to explore the question of transient disturbances when energizing the potline by the method employed at Mead. With the voltage regulators in normal position and the firing circuits energized, load was picked up by closing the a-c circuit breaker. There was no evidence of excessive transient currents or voltages. Figure 11, the oscillogram for test 16, is typical.

Klydonograph Records

In each test involving a-c circuit breaker operation, klydonograph elements were connected from line to ground on each phase of the transformer primary and between the terminals of one secondary wye group and the neutral of the opposite wye on the transformer of rectifier *13-A*.

In the tests involving cathode breaker operation, klydonograph elements were connected between the terminals of both secondary wye groups and the neutral of the opposite wye on the transformer of rectifier *13-A*. As pointed out previously, this is the most vulnerable part of the transformer winding to voltage surges. No indication of voltages as high as 2,000 volts, the minimum voltage which the klydonograph will record, were obtained in any test.

A cathode-ray oscilloscope was available, but intense magnetic fields present in the plant made its use impractical. Actually, klydonograph records having indicated no surge, cathode-ray records would have added little, if anything, to the test data.

Potline Resistance and Inductance

Because of the necessity of having reasonably accurate values of resistance and inductance for the calculation of short-circuit currents and surge voltages, these characteristics of the potline circuit were determined from the oscillograms made during these tests.

Calculations based on test oscillograms gave an average value of 0.0085 ohm for potline equivalent resistance, determined by dividing the difference between potline voltage under load and the polarization volts by the potline current.

Assuming the decrement of potline current to be an exponential, the time constant, L/R, is the time to the intersection of the initial tangent to the current curve and the current zero line. Determined this way, the average value of the time constant was 0.0419 second.

The inductance of the circuit is the product of time constant and resistance

$$\text{Inductance} = (L/R)R$$

Using the average values determined here, the inductance of the potline is 0.358 millihenry.

Rate of Rise of Arc-Back Current

In test 17, shown in Figure 8, an arc back occurred. The rate of rise for the first tube to arc back, measured by a tangent line, is 6.81 amperes per microsecond. The tube next conducting on the same wye also arced back following its conduction period the measured rate of rise in this case being 8.32 amperes per microsecond. This increase might be expected if it is assumed that saturation of the interphase transformer requires some interval of time, or magnitude of current, and hence exerts a modifying influence on the first arc back.

In spite of the high rate of rise, the maximum secondary phase currents were 20,580 and 20,190 amperes respectively for the first and second tubes to arc back. Since this arc back occurred after the ignitors were short-circuited and the record indicates that two rectifier units (*14-A* and *15-A*) did not contribute to the backfeed, the maximum possible arc-back current at this plant is undoubtedly slightly higher than is indicated on the oscillogram.

Conclusions

The primary objects of these tests were to determine the best operating procedure to be used in dropping potline load in proposed higher capacity ignitron rectifier installations and to establish whether or not measures should be taken to protect existing installations against equipment failure from switching surges. The following conclusions have been drawn in the light of these tests.

1. The best method of dropping load is by a-c circuit breaker operation with auxiliary power maintained. Using this method, no serious transient disturbances were observed in either secondary phase or cathode currents. Tripping the primary power supply removed any tendency to unbalance currents between parallel rectifiers, and the current in the transformer secondary windings was confined to two windings 180 degrees out of phase.

2. Load interruption by tripping the a-c circuit breaker and auxiliary power simultaneously is satisfactory at Mead. Observed transients in phase and cathode currents were not dangerously high, but were more severe than when auxiliary power was maintained while load was dropped, leading to the conclusion that provision should be made in proposed higher capacity installations to maintain auxiliary power. However, the tests show there is no cause for concern in cases of power failure in which the auxiliary power supply would be lost.

3. Load interruption by short-circuiting the firing circuits should not be considered in large rectifier installations because of excessive overcurrents and accompanying arc backs.

4. Load interruption by operation of the cathode circuit breakers is not recommended because of the stringent requirement on the last breaker to clear; this particular breaker must interrupt approximately one-half the normal potline current.

5. At Mead, it is satisfactory to pick up load by removing the short circuits on the firing circuits followed by closing the a-c circuit breaker. In general, it is recommended that load be picked up by removing the short circuits on the ignitor circuits after the a-c circuit breaker is closed. Such procedure reduces the maximum transient current on the a-c system since load current does not flow simultaneously with transformer inrush current.

6. The order of magnitude of inductance on aluminum potlines is less than 1 millihenry. This figure, originally arrived at by calculation, was confirmed by these tests.

7. It is improbable that failures of rectifier transformers on earlier installations were caused by voltage surges resulting from circuit-breaker operation.

References

1. Mercury-Arc Power Converters in North America, AIEE Committee Report. *AIEE Transactions*, volume 67, part II, 1948, pages 1031–59.

2. Load Pickup by a Group of Ignitron Rectifiers, L. W. Morton, D. I. Bohn. *AIEE Transactions (Electrical Engineering)*, volume 62, November 1943, pages 679–84.

3. Arc-Backs in Rectifier Circuits—Artificial Arc-Back Tests, R. D. Evans, A. J. Maslin. *AIEE Transactions (Electrical Engineering)*, volume 64, June 1945, pages 303–11.

4. Basis for Dielectric Tests for Rectifier Equipment, C. C. Herskind. *AIEE Transactions*, volume 70, part II, 1951, pages 1930–33.

5. Excitation Circuits for Ignitron Rectifiers, H. C. Myers, J. H. Cox. *AIEE Transactions (Electrical Engineering)*, volume 60, October 1941, pages 943–48.

Discussion

C. C. Herskind (General Electric Company, Schenectady, N. Y.): This paper is of interest for two reasons: first, because the increasing size of rectifier installations will accentuate the problems of load interruption, and second, because of unexplained transformer failures.

Power interruptions on rectifier installations may occur under any of the following four conditions:

1. Involuntary loss of a-c power. Switching operations in the a-c system

may interrupt the a-c supply to either the rectifier transformers or auxiliary power transformers or both. Even though duty on equipment is minimized by maintaining excitation, some interruptions without excitation probably will occur. Also, a wide variety of abnormal voltage and frequency conditions on the a-c system will be encountered just prior to interruption.

2. Involuntary loss of rectifier substation. This may arise from rectifier arc back and progressive tripping of rectifier units on overload.
3. Voluntary emergency trip from cell room.
4. Voluntary manual trip by rectifier operator.

The voluntary trips may be made by tripping a-c circuit breakers or cathode breakers, as appears preferable for any electrochemical installation having appreciable cell line inductance. However, it is not sufficient to consider only the voluntary interruptions in selecting a system of control.

As the size of rectifier installations is increased with a larger number of rectifier units feeding a single potline, load-dropping by means of the cathode breakers will impose greater duty on the rectifier unit which clears last. Not only will the cathode breaker be required to interrupt higher currents but the rectifier tanks will also carry higher currents during the last cycle or two. If these currents exceed the fault current capacity, the rectifier will arc back, imposing greater duty on the transformer in turn. Perhaps the duty during load-dropping may be reduced by co-ordinated action of phase control and breaker trip.

The differences in circuit action obtained with the different methods of load-dropping are clearly described in the paper. When load is dropped by opening the a-c breaker with excitation maintained, the potline current may take any path through the d-c windings of the rectifier transformers, as all anodes are free to conduct. Current flow is established in the anodes connected to windings 180 degrees apart as this gives the lowest inductance. This path presents only the leakage reactance between interlaced windings of the rectifier transformer. Any other path would introduce transformer magnetizing reactance. In the absence of any alternating voltages, the potline current decays smoothly to zero. When load is dropped by blocking ignitors, conduction is locked in those anodes already conductive at the instant of blocking. The average output voltage of the rectifier is zero but the alternating voltages in the transformer windings alternately aid and oppose the flow of potline current through these anodes, thereby causing pulsating currents. In the process the potline current is shifted around in an erratic manner, causing extinction of some anodes and accumulation of all current in the remainder. The process continues as the potline current decays until only one anode is conductive just before extinction.

Rectifier transformer failures similar to that shown in Figure 2 of the paper have occurred at different installations. The nature of the damage indicates that the failures were caused by voltage stress. However, to the best of our knowledge a satisfactory explanation has not yet been developed.

A review of the situations in which failures of this type have occurred on General Electric transformers indicates that the following conditions apply:

1. All failures have occurred on units which are heavily loaded.
2. All failures have occurred on assemblies having only one tank connected to each transformer winding.
3. All failures have occurred on potlines with less than 50,000 amperes rated capacity.
4. Two types of transformers have failed, that is, both zigzag double-wye connected types.
5. Arc backs do not always occur at the time of failure.

In an effort to determine the cause of these transformer failures, extensive studies and tests have been made at the factory on transformer windings and surge absorbers. These tests have not disclosed any weaknesses. In fact, voltage tests have shown that impulse voltages of the order of 50 kv are required to cause breakdown between interlaced windings. These voltages far exceed those expected in service. However, with the hope of reducing the probability of failure, the insulation between interlaced windings has been increased and the thyrite surge absorbers have been modified so as to improve their effectiveness.

Limited tests have also been made in the field with the installation of film-type surge recorders on a few rectifier units. No surge voltages were indicated on the records returned from the field.

Comprehensive field tests for the purpose of determining the cause of failure have been proposed. Lacking a hypothesis regarding the mechanism of failure, the prospect of obtaining an answer appears unfavorable. Further deterents to elaborate field tests of the type required are the hazards to equipment involved in the test and the high cost of tests. In lieu of such tests, the painstaking search for clues as to the cause of failure must be continued.

Lysle W. Morton (General Electric Company, Schenectady, N. Y.): The paper is of outstanding interest to rectifier application and design engineers of my company. The subject is a timely one, for it relates to the perplexing problem of voltage surges in large rectifier installations that have caused rectifier transformer failures.

We have had experience similar to those described by the authors, when rectifier transformers failed. In certain cases the failures appear to be caused by transient voltage surges, while in other cases involving rectifier units of early vintage, the failures obviously resulted from mechanical stresses (evidently repeated arc backs). Failures attributable to mechanical stresses have been eliminated by special bracing and clamping of windings in rectifier transformers and use of high-speed anode switchgear. The failures apparently caused by transient overvoltages are guarded against by improved surge protection and increased dielectric strength of the winding insulation.

HIGH-VOLTAGE SURGES

The source of the transient overvoltages has been the subject of much study and exploratory testing in our organization, but no convincing answer to the problem has been found. We infer from the authors' statement of the purposes of tests made at Mead that a similar search for the source of transient overvoltages had an important bearing on the investigation.

One widely held theory was that the rectifier transformers having zigzag windings were particularly susceptible to failure of the transient overvoltage type. This idea has been disproved because failures have also occurred involving rectifier transformers without zigzag windings.

There has been a persistent theory that the source of the overvoltage may be transient disturbances associated with load-dropping of either planned or accidental origin. It was for this particular reason that we studied the paper with such great interest. We regret, as probably the authors also do, that the Mead tests did not find the source of these overvoltages. It is believed that there is enough evidence to establish that such overvoltages are generated at rare intervals and do exist. We agree with the authors' conclusion 7. We believe that the search for the true source of high voltage should continue.

THE BEST METHOD OF LOAD-DROPPING

With regard to the other conclusions of the paper, there are a few additional remarks that should be made. It is our opinion that the tests made at Mead by the authors brought to light an interesting, new, and workable method of load-dropping for large rectifier installations. The authors have concluded that load-dropping "by a-c circuit-breaker operation with auxiliary power maintained" is best.

The question of whether the new method or the older widely used method of "load interruption by operation of cathode breakers" is best has many pros and cons. Experience has indicated that simultaneous operation of the cathode breakers has, in fact, not been an excessively severe duty on the last breaker to open.

Load-dropping tests made at Mead in 1942 and reported in Figure 6 of reference 2 of the paper agree with these more recent Mead tests that the last breaker interrupts approximately one-half the potline current. But, such values of current are not too high for the type of breakers usually employed as cathode breakers in these large installations. Any breaker suitable for this type of service which finds itself in distress when interrupting overcurrents equal to one-half potline current ought to have attention from the designers. The practice of dropping load by cathode-breaker operation has been employed for more than 10 years in many large rectifier installations without excessive breaker maintenance, according to reports I have received.

Furthermore, it is possible that some users would rather take such maintenance as may be necessary on the cathode breakers than on main a-c power circuit breakers.

There is also a specific objection to the newly proposed method. When the rectifier excitation and auxiliary power is maintained during main a-c breaker operation, the auxiliary power must be supplied from the a-c system from a connection ahead of the main a-c breaker feeding the rectifier units. Tapping auxiliary power at this point complicates the substation layout. In some cases it may even be necessary to add an expensive heavy-duty a-c breaker to control the auxiliary power or else risk hazard to personnel from alternative inadequate protective means.

It is our conclusion that there are now two workable methods for dropping load in large rectifier installations. One may safely make his choice based on his own local requirements and personal preferences.

H. Winograd (Allis-Chalmers Manufacturing Company, Milwaukee, Wis.): The paper is timely in view of the trend towards higher aluminum potline currents and voltages. It adds useful information to what has been published previously on switching aluminum potline circuits.

Load-switching tests were conducted in March 1939, at the Alcoa plant of the Aluminum Company of America, in the first rectifier installation for aluminum reduction[1] and described by the writer.[2] The tests were made for the purpose of developing a load-switching method without using a potline circuit breaker. As a result of these tests, a potline switching method was adopted using the firing control of the rectifiers (by grid control) for switching on the load and gang-tripping of the cathode circuit breakers for switching it off.

This switching method was applied in the second rectifier installation at Alcoa for a 55,000-ampere 625-volt potline, which was put in operation in May 1940. That was also the first installation where high-speed anode breakers and semihigh-speed cathode breakers were employed, and the a-c line breakers of individual rectifier units were omitted. The switching and protection system of this installation set the pattern for most of the subsequent rectifier installations in aluminum reduction plants and other high-current electrochemical applications. Switching a potline with an a-c line breaker was not feasible because two or more potlines were supplied from a common a-c bus.

Switching off the load by gang-tripping of the cathode breakers is used for many 60,000-ampere potlines operating at 650 to 750 volts. The writer has not heard of any difficulties encountered with this method. In installations with which the writer was connected, the opening time of the individual cathode breakers was adjusted with the aid of an electronic timer, so that the breakers supplying a load circuit opened simultaneously with a time difference of less than 1/3rd cycle.[2] This helped to distribute the interrupting duty among the breakers. Even if the initial adjustment is not maintained, the same breaker is not likely to be subjected to the full circuit interrupting duty repeatedly. However, one breaker should be capable of interrupting the circuit.

A voltage surge resulting from an interruption of the load would appear across the potline circuit, because that is where the major part of the inductance is located. It might have some effect on the rectifier transformer by shifting the negative bus potential to ground in the positive direction. However, the leakage current through the ground between pots which are at different potentials in the potline circuit would tend to reduce the magnitude of the surge.

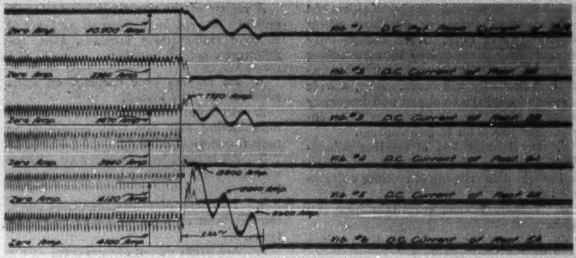


Figure 1. Oscillogram of currents during load-dropping test on aluminum potline by grid blocking of rectifiers

The interrupting duty imposed by a high-current potline on a d-c breaker cannot be minimized. This was pointed up vividly by a test made in March 1940 in the second Alcoa rectifier installation before it was put in service. One of the cathode breakers was tested in the potline circuit, with the pots short-circuited and with an external series resistor for adjusting the current. The potline consisted of four potrooms in series, and the circuit inductance was 0.7 to 0.8 millihenry. The current interrupted by the breaker was increased in steps. On the third test, the breaker interrupted 37,000 amperes. The arcing time was 4.5 cycles, and the fumes from the arc hit the ceiling, about 15 feet above the breaker. The breakers were subsequently provided with larger arc chutes and stronger blowout fields and performed satisfactorily.

With the increase in potline ratings to 100,000 or 125,000 amperes, approximately double that of the present lines, the energy stored in the circuit inductance would be about four times as high. It is unlikely that any existing d-c breaker would be capable of interrupting this circuit successfully. Because of this limitation, the use of an a-c line breaker for switching off the load of these high-current potlines is the best solution available at the present time. With this method, however, some of the operating flexibility of the present installations might be lost.

Figure 1 of the discussion is an oscillogram obtained during the load switching tests at Alcoa.[2] It shows the potline and rectifier currents when the load was switched off by applying grid blocking to all the rectifiers simultaneously. The potline was supplied by five double-rectifier units constituting a 30-phase system. The anodes which were not conducting when blocking was applied remained blocked. The currents of some conducting anodes were transferred to other conducting anodes in the same or another rectifier by commutation between phases. The current transfer between rectifiers was facilitated by their phase displacement in the 30-phase system. The potline current was finally conducted by one phase in each of two units. The potline inductance maintained the current flow over the entire voltage cycle for several cycles until the stored energy was dissipated and the currents reached zero. The 60-cycle components of the currents were produced by the phase voltages. This switching method was not adopted because the overloading of one phase and the anodes connected to it for several cycles was considered undesirable. One phase of one of the units might conduct the entire potline current during its decay.

This test was equivalent to tests 14 and 17 of the paper, in which blocking was applied to the rectifiers by short-circuiting the ignitors. The more irregular decline of the currents in Figure 7 and 8 of the paper was caused by arc backs and reignitions after the ignitors were short-circuited. In Figure 7 there were arc backs in rectifier *17A* and reignitions in rectifiers *13A* and *18A*. In Figure 8 there was an arc back in *13A* and reignitions in *18A*.

The authors should be cautioned not to use the arc-back currents measured during these tests for estimating the magnitude of arc-back currents obtained in normal service. The latter would be considerably higher, because the other phases of the faulted rectifier as well as the other rectifiers would be free to feed current to the faulted anode.

REFERENCES

1. ALCOA RECTIFIER INSTALLATION, J. Elmer Housley, H. Winograd. *AIEE Transactions,* volume 60, 1941, pages 1266–74.

2. Discussion by H. Winograd of reference 2 of the paper, pages 964–65.

S. J. Pope, J. K. Dillard, and **C. R. Marcum:** We wish to express our appreciation to the discussers for the interesting points which have been raised about our paper.

Mr. Herskind indicated that in his experience all rectifier transformer insulation failures had occurred on units where only one tube was connected to each transformer winding and where the potline current rating was less than 50,000 amperes. In our experience most failures have occurred where two tubes were connected to each transformer winding and, while most of the failures have been on potlines with less than 50,000 amperes rating, others have occurred on potlines rated 60,000 amperes.

We also have had surge indicators on certain units in the field. Some of these indicators have been connected for as long as 9 years without any surge indication.

While it is true that transient disturbances in the a-c system may exist just prior to involuntary interruptions of potline load, the excitation circuits are designed to operate down to 50 per cent of applied alternating voltage and with some variation in frequency. There will be some shifting of the phase of the firing peak but this is not important after the alternating voltage is removed.

Mr. Morton points out that master-tripping of the cathode breakers has proved satisfactory where the potline current is 60,000 amperes or less. For the new potlines with rating up to 125,000 amperes the last breaker out may not be capable of interrupting one-half of the potline rated current. Such severe duty is not imposed on the a-c breaker.

Mr. Morton stated that in some cases an additional a-c circuit breaker may be required to obtain the auxiliary power from the system ahead of the a-c breaker. In nearly all large potlines, however, there is other electric equipment in the plant requiring a separate feeder which may be used for rectifier auxiliary power for short periods. In these cases the additional breakers may be low voltage and therefore relatively inexpensive.

Tests made in the laboratory indicate that the excitation circuits will continue to ignite the tubes for approximately 1/2 cycle after the alternating voltage is removed. If this time could be extended to 1 cycle by excitation circuit design, auxiliary power could be obtained from the load side of the a-c breaker and all units would share the current during the load-dropping period. With present excitation circuits, at least half of the units will conduct on two phases which are 180 degrees apart and thereby reduce the potline current to zero without any disturbances.

Our thanks go to Mr. Winograd for confirming a number of points in our paper. We concede that the arc backs shown on the oscillograms of the test were affected by load shifting between tubes and units and by the decay of the potline current. The paper acknowledged that a normal arc back would result in a higher reverse current than indicated by the oscillogram.

All of us seem to be in agreement that work should be continued to locate the source of voltage breakdown failures of rectifier transformers. As larger potlines are put into service, our prediction is that the a-c breaker method of load-dropping will be the accepted practice as it has been at the Mead Plant for some time.

Mechanized Billing of AMA Toll Messages

FRANK D. SLADE
NONMEMBER AIEE

THE billing of 140,000,000 toll messages a month to 17,000,000 toll using customers, representing about 60 per cent of the total telephone customers, is a major operation involving approximately 15,000 Bell System clerical employees. Toll message volumes have more than doubled in the past 10 years, and further growth seems likely. While many important improvements have been made in the billing processes in order to handle the increased volume more effectively, the mechanization of this work continues to remain a challenge.

The introduction of automatic message accounting (AMA) in the Bell System in 1948 presented an opportunity to mechanize some of the operations involved in toll billing, especially the sorting of AMA toll messages into telephone number order and the timing of such messages. The initial AMA system was described in 1950,[1] and a new development in the AMA system, the tape-to-card converter which produces a punched card for each toll message, is described in a companion paper.[2]

Although the number of toll messages now recorded on AMA tapes represents only about 3 per cent of the total toll messages, this percentage is expected to double by the end of 1953 and continued rapid expansion of AMA toll message recording is anticipated.

This paper describes the operations followed by the New Jersey Bell Telephone Company in billing toll messages recorded on AMA equipment. The operations involved in billing long-distance toll messages dialed on a trial basis by customers in Englewood and Teaneck, N. J., are not covered, although the same card is used for billing regular AMA toll messages and long-distance messages. The New Jersey Company at present accounts for about 80 per cent of the total AMA recorded toll messages and is the only Bell System company currently using punched cards for toll message billing. The Illinois Bell Telephone Company and the New York Telephone Company are scheduled to adopt punched card toll billing for AMA messages early this year, and three other Bell companies are planning to introduce this type of billing later in 1953. AMA toll messages billed manually are handled in about the same manner as operator-ticketed toll messages.

For many years experiments have been conducted in an effort to find a mechanical method of rating, sorting, and billing toll messages which would provide economies over the highly efficient manual procedures that had been developed over a period of many years. In the fall of 1948, the New York Telephone Company began a trial of punched card toll billing in its Lexington Accounting office in New York City wherein the significant data shown on toll tickets, prepared in the usual manner by operators, were transferred to punched cards, using a specially designed key punch with an automatic ticket feed. While more than 20,000,000 toll messages were billed by means of punched cards during the period of the trial, the plan was discontinued in December 1950 after much experimentation, when it was concluded that the manual transcription of information from toll tickets to punched cards was uneconomical. One of the principal factors affecting the outcome of the trial was the difficulty encountered in getting the data for each message (generally of low billing value) punched into the cards accurately and at high speed.

With the introduction of the AMA tape-to-card converter, the accurate and rapid preparation of punched cards from original records became a reality, and mechanical billing of AMA toll messages became an economical operation where the volume of toll messages to be billed was at least 200,000 messages per month. While the procedures followed by the New Jersey Company are producing satisfactory results, the many machine and card handling operations involved create a complicated process, which, it is believed, will eventually evolve into a more efficient and simplified procedure.

Preparation of Toll Message Card

The punched card, Figure 1, as it emerges from the tape-to-card converter contains all of the basic information required for rating and billing. The follow-

---

Paper 53-114, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted November 19, 1952; made available for printing December 12, 1952.

Frank D. Slade is with the American Telephone and Telegraph Company, New York, N. Y.

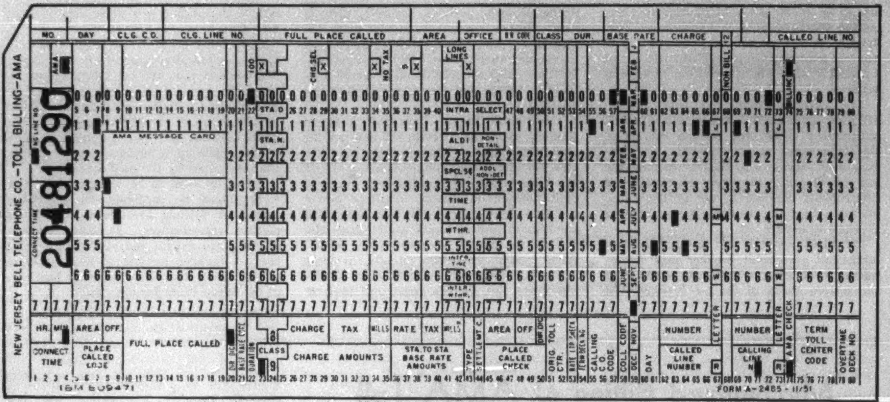

NEW JERSEY BELL TELEPHONE CO.—TOLL BILLING—AMA
NO.
DAY
CLG. C.O.
CLG. LINE NO.
FULL PLACE CALLED
AREA
OFFICE
RR CODE
CLASS
DUR.
BASE RATE
CHARGE
CALLED LINE NO.
AMA MESSAGE CARD
2048129
CONNECT TIME
AREA
OFF.
PLACE CALLED CODE
FULL PLACE CALLED
CHARGE
TAX
CHARGE AMOUNTS
RATE
STA. TO STA. BASE RATE AMOUNTS
PLACE CALLED CHECK
ORIG. TOLL CTR.
CALLING C.O. CODE
COLL CODE
DAY
NUMBER
CALLED LINE NUMBER
LETTER
CALLING LINE N
AMA CHECK
TERM TOLL CENTER CODE
OVERTIME DECK NO.
NON BILL
BILLING
IBM 609471
FORM A-2485 - 11/51

Figure 1. Toll message card as produced by converter

ing data are punched into the cards:

Column	Punched Data
1–4	connect time of message (using 24-hour clock)
4	special X punch to indicate an AMA message
5–7	called area code (column 7 only for a regular AMA message)
8, 9, 20	called central office code
22, 23	duration of conversation in chargeable minutes
55–57	calling central office code
58	month code (for collator date sequence check)
59–61	month and date of message
62–66	called telephone number
67	party letter of called telephone number
69–72	calling telephone number
74	AMA check ("9" punch); indicates all data were received by punch unit of converter from the input tape
74	billing X indicates a billable message

After verifying the count of message cards punched by the converter with the number of messages recorded on the AMA computer registers, the cards are replaced in the regular cardboard boxes used for shipping and are forwarded three times a day to the punched card accounting offices in Newark and Elizabeth, which are, respectively, 1 and 6 miles from the AMA Accounting Center.

The Newark punched card accounting office processes about 1,700,000 toll messages per month and prepares toll service statements for 73,000 customers whose telephone bills are prepared at Newark and Paterson. The Elizabeth punched card accounting office processes about 1,400,000 toll messages per month and prepares toll service statements for 75,000 customers whose telephone bills are prepared at Elizabeth and Pennsauken, near Camden. As a temporary arrangement, in order to produce a more even work load at the two punched card accounting offices, toll statements are also prepared at Elizabeth for a small number of customers billed at Paterson.

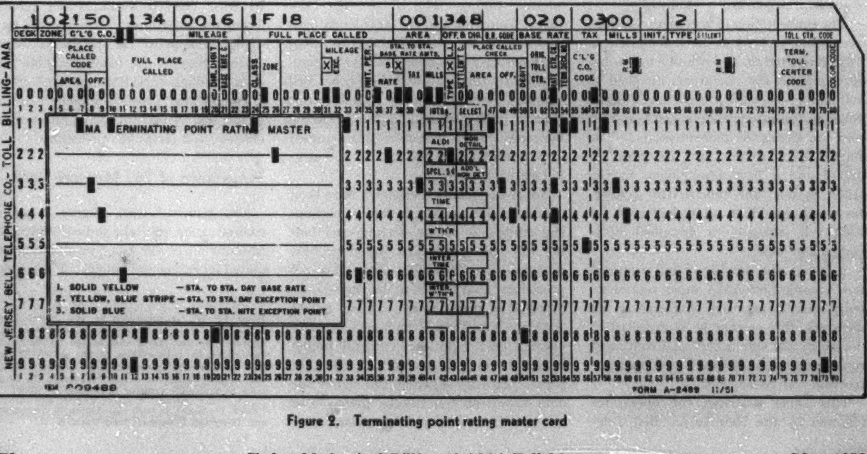

NEW JERSEY BELL TELEPHONE CO.—TOLL BILLING—AMA
102150 134 0016 1F18 001348 020 0300 2
DECK
ZONE
C'L'G C.O.
MILEAGE
FULL PLACE CALLED
AREA
OFF. & DIG.
R.R. CODE
BASE RATE
TAX
MILLS
INIT.
TYPE
TOLL CTR. CODE
PLACE CALLED CODE
FULL PLACE CALLED
AREA
OFF.
CLASS
STA. TO STA. BASE RATE AMTS.
PLACE CALLED CHECK
C'L'G C.O. CODE
TERM. TOLL CENTER CODE
COLOR CODE
AMA TERMINATING POINT RATING MASTER
1. SOLID YELLOW —STA. TO STA. DAY BASE RATE
2. YELLOW, BLUE STRIPE —STA. TO STA. DAY EXCEPTION POINT
3. SOLID BLUE —STA. TO STA. NITE EXCEPTION POINT
FORM A-2489 11/51

Figure 2. Terminating point rating master card

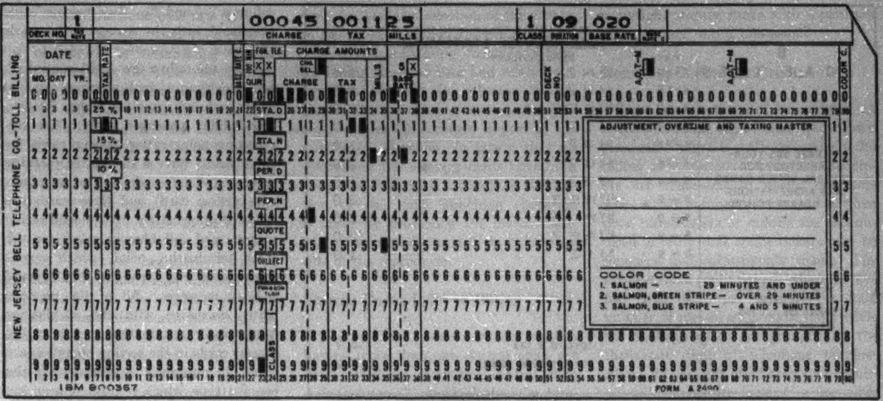

NEW JERSEY BELL TELEPHONE CO.-TOLL BILLING
DECK NO.
CHARGE
TAX
MILLS
CLASS
DURATION
BASE RATE
DATE
MO.
DAY
YR.
TAX RATE
25 %
15 %
10 %
CHARGE AMOUNTS
STA. D
STA. N
PER. D
PER. N
QUOTE
COLLECT
CLASS
A.D.T.-M
ADJUSTMENT, OVERTIME AND TAXING MASTER
COLOR CODE
1. SALMON — 29 MINUTES AND UNDER
2. SALMON, GREEN STRIPE — OVER 29 MINUTES
3. SALMON, BLUE STRIPE — 4 AND 5 MINUTES
IBM 800357
FORM A 2490

Figure 3. Duration rating master card

Punched Card Machines

The four principal types of punched card machines used in billing AMA toll messages are as follows:

Sorters

Sort cards, one column at a time, into one of 12 numbered pockets or a reject pocket.

Count number of cards sorted.

Rated speed is 650 cards per minute.

Collators

Compare cards for ascending numerical sequence, selecting those out of numeric order.

Compare up to 16 columns in one card with matching columns in following cards, rejecting nonmatching cards.

Rated speed is 240 cards per minute for single-feed, or 480 per minute for double-feed.

Reproducing Punches

Gang punch data from one card into another card.

Detect a double punch or a blank column where one punch should be present.

Compare all columns of one card with the following cards.

Duplicate punches in one card to another card.

Rated speed is 100 cards per minute.

Accountings Machines

Print up to 88 characters on one horizontal line simultaneously.

Accumulate totals into counters and print totals upon recognizing a control punch.

Automatically position new pages of forms to be printed.

Rated speed is 100 cards per minute.

The uses made of these machines in the rating and billing of AMA toll messages are described in succeeding paragraphs.

Primary Rating Operation

The boxes of message cards, upon being received in the punched card accounting office, are examined to insure that all of the boxes for the recorder groups involved have been received, it being the established practice for the AMA accounting office to release all of the cards for a

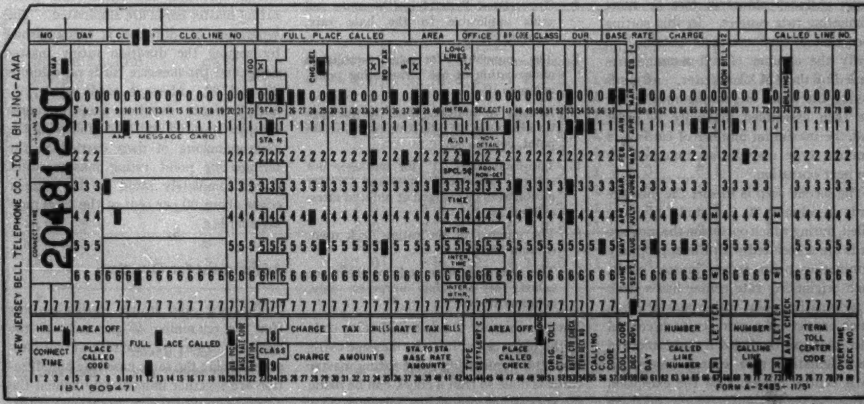
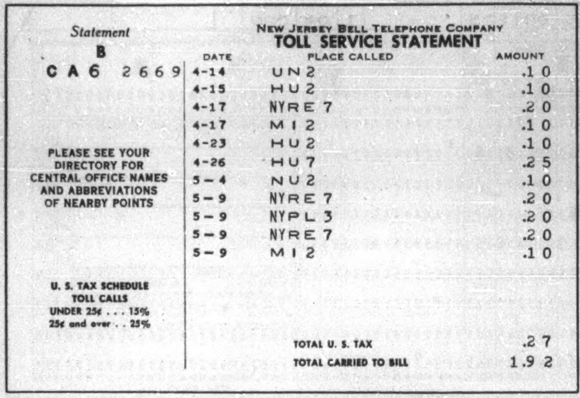

NEW JERSEY BELL TELEPHONE CO.-TOLL BILLING-AMA
MO.
DAY
CL
CLG. LINE NO.
FULL PLACE CALLED
AREA
OFFICE
CLASS
DUR.
BASE RATE
CHARGE
CALLED LINE NO.
2048 1290
CONNECT TIME
AMA MESSAGE CARD
STA. D
STA. N
CHG. SEL.
NO TAX
LONG LINES
INTRA
SELECT
A.D.I.
SPCL. 56
TIME
WTHR.
INTER. TIME
INTER. WTHR.
NON BILL 12
BILLING
AMA CHECK
LETTER
HR.
MIN.
AREA
OFF.
CONNECT TIME
PLACE CALLED CODE
FULL PLACE CALLED
CLASS
CHARGE
TAX
MILLS
RATE
CHARGE AMOUNTS
STA. TO STA. BASE RATE AMOUNTS
TYPE
SETTLEMENT
PLACE CALLED CHECK
ORIG. TOLL CTR.
CALLING C.O. CODE
COLL. CODE
DAY
NUMBER
CALLED LINE NUMBER
CALLING LINE
TERM TOLL CENTER CODE
OVERTIME DECK NO.
IBM 809471
FORM A-2485-11/51

Figure 4. Toll message card ready for billing



Statement
B
CA6 2669

NEW JERSEY BELL TELEPHONE COMPANY
TOLL SERVICE STATEMENT

DATE	PLACE CALLED	AMOUNT
4-14	UN 2	.10
4-15	HU 2	.10
4-17	NYRE 7	.20
4-17	MI 2	.10
4-23	HU 2	.10
4-26	HU 7	.25
5-4	NU 2	.10
5-9	NYRE 7	.20
5-9	NYPL 3	.20
5-9	NYRE 7	.20
5-9	MI 2	.10
	TOTAL U. S. TAX	.27
	TOTAL CARRIED TO BILL	1.92

PLEASE SEE YOUR
DIRECTORY FOR
CENTRAL OFFICE NAMES
AND ABBREVIATIONS
OF NEARBY POINTS

U. S. TAX SCHEDULE
TOLL CALLS
UNDER 25¢ . . . 15%
25¢ and over . . 25%

Figure 5. Customer's portion of toll service statement

recorder group at one time. The individual boxes are marked to show the number of the box and the total number of boxes for the recorder group.

A recorder group may consist of a single central office, such as Cranford 6, or it may comprise two or more offices, such as Leonia 4 and Fort Lee 8, calls for which are perforated on common recorders. The central offices in a recorder group may or may not be in the same originating rate center. Since the messages in a recorder group are not separated by central office when being processed through the tape-to-card converter, the first step in the rating operation is to sort the cards on column 57 by originating central office code, if the recorder group includes offices in two or more originating rate centers. In this sorting operation a card count is obtained to verify the number of toll messages recorded by the AMA computer. All cards having the same originating rate center are then assembled manually in order that the maximum number of cards of a common type may be accumulated for the next operation.

The second step is to sort the message cards on columns 20, 9, 8, and 7 by place called, during which operation the message cards are associated with the appropriate terminating point rating master cards. Each originating rate center has a deck of terminating point rating master cards which consists of a card for each central office that can be dialed by customers of the originating rate center. Since all AMA central offices, with the exception of those around Camden, in the southern part of the state, have the same dialing area, the number of cards in the terminating point rating master deck for each originating rate center is approximately the same. The offices in and around Camden have a much smaller calling range and therefore the terminating point rating master decks consist of fewer cards. Currently, the rating decks for northern New Jersey offices consist of about 450 cards and for offices in the southern part of the state, about 125 cards.

The terminating point rating master cards, Figure 2, which are yellow in color, show the following information which is punched into the message cards:

1. The initial period station-to-station day rate (commonly referred to as the base rate), columns 36–38.
2. The related Federal tax in cents and mills applicable to the base rate, columns 39–42.
3. The punching for the established abbreviation of the terminating point, columns 10–13.
4. A code to indicate that the message terminates at an interstate or intrastate point, column 43.
5. The class code, "1" for messages not subject to reduced rates, and "2" for messages to be charged at reduced rates, column 24.
6. The number of the rating deck used, column 54.
7. The area and office code of the place called for checking purposes only, columns 47–50.

The punching of the area and office code of the place called into the message cards from the terminating point rating master cards is done so that a check can be made by comparing the newly punched data in columns 47–50 with the punching in columns 7–9 and 20. Any message cards associated with the wrong master cards will cause the reproducer to stop.

There are only a few AMA central offices which can originate messages to points where reduced rates for night, Sunday, and holiday messages apply. The terminating point rating master cards for these points have a distinctive marking (blue stripe across the top of an otherwise yellow card) and these cards are followed immediately by a second rating master card (all blue in color) for the same terminating point, which carries the reduced rate for night, Sunday, and holiday messages. All of the message cards associated with the all-blue or blue-striped terminating point rating master cards in the sorting operation are manually examined to determine if the connect time which is printed near the left end of the card by the tape-to-card converter is later than 1759 and earlier than 0430, in which case reduced rates are applicable. Those cards subject to reduced rates are filed behind the all-blue master card and the others are filed behind the blue-striped master card.

After the message cards have been associated with their respective terminating point rating master cards, they are gang punched with the pertinent data carried in the master cards. During the punching operation, the machine verifies the accuracy of part of the punching. However, to insure that all columns of the message cards have been accurately punched, the cards are placed in the comparing side of the reproducer, which completes the verification.

The message cards are next sorted by duration (length of conversation), during which operation the terminating point rating master cards are eliminated. The separation by duration is accomplished by sorting the duration rating master cards and the message cards on columns 23 and 22. Message cards of 1-, 2-, and 3-minute duration are, of course completely rated and taxed following the gang punching of data carried in the terminating point rating master cards. Cards completely rated in this manner cover about 60 per cent of the total message cards.

Adjustment for Overtime

The next step in the rating process (for the remaining 40 per cent of the cards) is to sort by base rate so that messages having the same base rate and duration are together immediately behind the appropriate duration rating master cards. In this rating operation, all message

8

cards having the same base rate may be processed at the same time regardless of the originating or terminating central office, since it is the base rate charge, length of conversation, and class of message only that determine the amount to be billed. The duration rating master decks include cards for each minute of duration from 4 to 99 minutes for each base rate charge of 10, 15, and so forth, to 45 cents, inclusive; see Figure 3. In addition, there are special cards in file for rating messages to time and weather service. Messages of more than 99 minutes' duration are rare and are specially handled in the AMA computer, which records the pertinent data for such messages on a "straddle" tape which is processed through the AMA printer and then manually handled.

Following these sorting operations and before gang punching the charge data into the message cards, the duration rating master cards and message cards are processed through the collator to verify that the message cards have been correctly associated with the duration rating master cards. In this operation, a comparison is made of the punching in columns 22 and 23 (duration) 36, 37, and 38 (base rate) of the duration master and message cards. At the same time, any improperly associated message cards are mechanically ejected and after being examined are associated with their proper duration rating master cards.

After the association check has been completed, the charge and tax amounts are gang punched into the message cards in columns 25–35. In addition, an *X* punch is made in column 29, which controls the tabulator selection of the charge to be billed. In the absence of an *X* punch, the accounting machine selects for billing the base rate and related tax amounts punched in columns 36–42. During the punching operation, the machine verifies the accuracy of part of the punching. However, to insure that all columns of the message cards have been accurately punched, a complete verification of the gang punching operation is made by processing the master cards and message cards through the comparing side of the reproducer.

Storing Cards Prior to Billing

The message cards for rate centers having more than one central office are then sorted by originating central office code (column 57), at which time the duration rating master cards are eliminated. The cards for each central office are then sorted by the thousands digit of telephone number (column 69). During this sort, the duration master cards for single office rate centers are withdrawn. In both of these operations the message cards of 1-minute duration are kept separate for subsequent handling in the application of credit tickets.

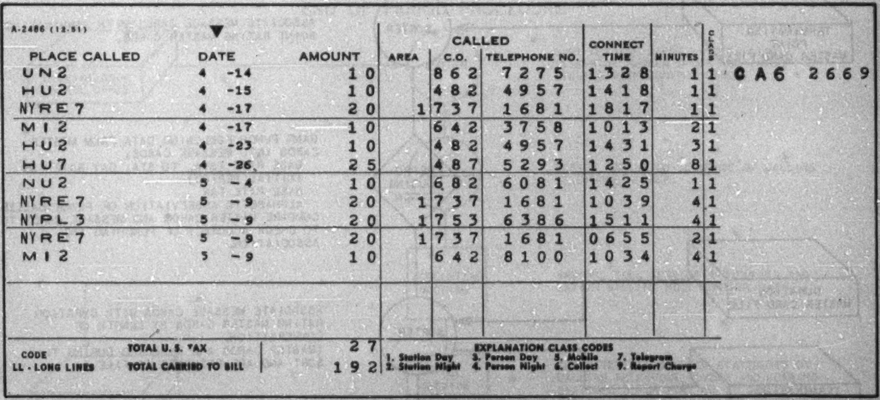
Before the message cards are placed in file by central offices, they are needle-checked to detect possible discrepancies in date order resulting from (a) the fact that the AMA business day ends at 3 a.m. or (b) any mishandling of cards. At this point, the punching of the message cards has been completed and they are ready for listing on toll service statements. The cards, Figure 4, are filed by thousands in upright storage racks with separators between each day's business.

Handling Claims for Credit

Some errors and claims for credit do, of course, require adjustment and every reasonable effort is made to prevent the billing of erroneous amounts. Customers are requested to notify operators when wrong numbers are reached or when poor transmission or cutoffs are encountered in order that credit tickets may be prepared and adjustments made. The application of credit tickets, which involve less than 0.005 per cent of the total number of messages, is made daily while the message cards are grouped by the thousands digit of telephone number. About 90 per cent of all messages requiring adjustment are located in the cards of 1-minute duration. The message cards involving credits are withdrawn from the billing process and are associated with the credit tickets for subsequent clerical handling.

Message cards involving credits are located by a block sorting process which requires the minimum handling of cards. The identification of a particular card is, of course, facilitated by the presence of the telephone number and connect time which are printed on the cards by the tape-to-card converter.

Most wrong number credit tickets result from incorrect dialing by customers, although some are caused by incorrect handling at terminating manual switchboards. No attempt is made to locate message cards for wrong number credit tickets in the 2-minute and longer



A-2486 (12-51)

PLACE CALLED	DATE ▼	AMOUNT	CALLED AREA	C.O.	TELEPHONE NO.	CONNECT TIME	MINUTES	CLASS	
UN2	4 -14	10		862	7275	1328	1	1	CA6 2669
HU2	4 -15	10		482	4957	1418	1	1	
NYRE7	4 -17	20	1	737	1681	1817	1	1	
MI2	4 -17	10		642	3758	1013	2	1	
HU2	4 -23	10		482	4957	1431	3	1	
HU7	4 -26	25		487	5293	1250	8	1	
NU2	5 -4	10		682	6081	1425	1	1	
NYRE7	5 -9	20	1	737	1681	1039	4	1	
NYPL3	5 -9	20	1	753	6386	1511	4	1	
NYRE7	5 -9	20	1	737	1681	0655	2	1	
MI2	5 -9	10		642	8100	1034	4	1	

CODE
LL - LONG LINES

TOTAL U. S. TAX 27
TOTAL CARRIED TO BILL 192

EXPLANATION CLASS CODES
1. Station Day
2. Station Night
3. Person Day
4. Person Night
5. Mobile
6. Collect
7. Telegram
9. Report Charge

Figure 6. Business office portion of toll service statement

duration cards, since it is assumed that a message of more than 1-minute duration did not involve a wrong number. In rare cases, messages of 2 minutes or more may involve valid wrong number claims and the adjustments are handled by the business office.

Message cards of 2-minute and longer duration are searched, however, in connection with the application of credit tickets for cutoffs or poor transmission. Message cards involved in such claims are removed and associated with the related credit tickets. In allowing credit for cutoffs, the combined time of the AMA message and the reconnection is examined. An allowance of 1 minute is made, and where adjustment is proper, a new card is prepared showing the correct duration and charge. Adjustment of charges on messages involving poor transmission is made in accordance with the agreement reached by the customer and the operator who handled the claim.

All credit tickets properly noted as to disposition are sorted into telephone number order at the close of the billing period and are forwarded to the business office for reference purposes.

Preparation of Toll Statement

To spread the work load as evenly as possible throughout the month, all Bell System companies use rotation or cycle billing. Under this plan, customers are billed once a month, with different groups of customers being billed at different fixed dates during the month. Since toll service statements are issued on the same basis, their preparation is an almost continuous process.

When the message cards for the last date of the billing cycle have been completed, the cards are removed from file and are sorted into telephone number order with date sequence being maintained. In this operation, billing book control cards are mechanically interfiled with the message cards at selected points so that control totals may be established by units. These unit totals are used in balancing the amounts subsequently entered on customers' bills. The telephone number sequence of all message cards is verified in the collator before the cards are listed on toll service statements.

The toll service statements used for listing AMA toll messages are continuous forms consisting of two parts, the customer's portion and the business office portion. The customer's portion, Figure 5, of the toll service statement shows the central office designation and telephone number, date, place called, charge (excluding tax) for each message, and the total tax and total amount (including tax) to be entered on the customer's bill.

The business office portion, Figure

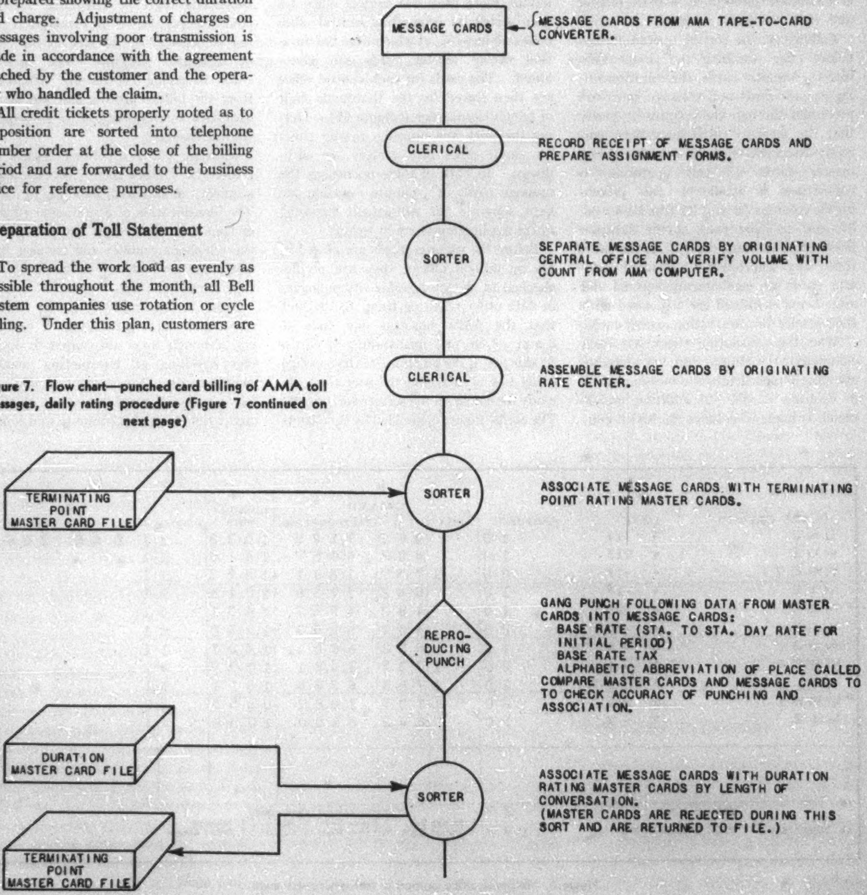

MESSAGE CARDS
MESSAGE CARDS FROM AMA TAPE-TO-CARD CONVERTER.
CLERICAL
RECORD RECEIPT OF MESSAGE CARDS AND PREPARE ASSIGNMENT FORMS.
SORTER
SEPARATE MESSAGE CARDS BY ORIGINATING CENTRAL OFFICE AND VERIFY VOLUME WITH COUNT FROM AMA COMPUTER.
CLERICAL
ASSEMBLE MESSAGE CARDS BY ORIGINATING RATE CENTER.
TERMINATING POINT MASTER CARD FILE
SORTER
ASSOCIATE MESSAGE CARDS WITH TERMINATING POINT RATING MASTER CARDS.
REPRO-DUCING PUNCH
GANG PUNCH FOLLOWING DATA FROM MASTER CARDS INTO MESSAGE CARDS:
BASE RATE (STA. TO STA. DAY RATE FOR INITIAL PERIOD)
BASE RATE TAX
ALPHABETIC ABBREVIATION OF PLACE CALLED
COMPARE MASTER CARDS AND MESSAGE CARDS TO TO CHECK ACCURACY OF PUNCHING AND ASSOCIATION.
DURATION MASTER CARD FILE
SORTER
ASSOCIATE MESSAGE CARDS WITH DURATION RATING MASTER CARDS BY LENGTH OF CONVERSATION.
(MASTER CARDS ARE REJECTED DURING THIS SORT AND ARE RETURNED TO FILE.)
TERMINATING POINT MASTER CARD FILE

Figure 7. Flow chart—punched card billing of AMA toll messages, daily rating procedure (Figure 7 continued on next page)

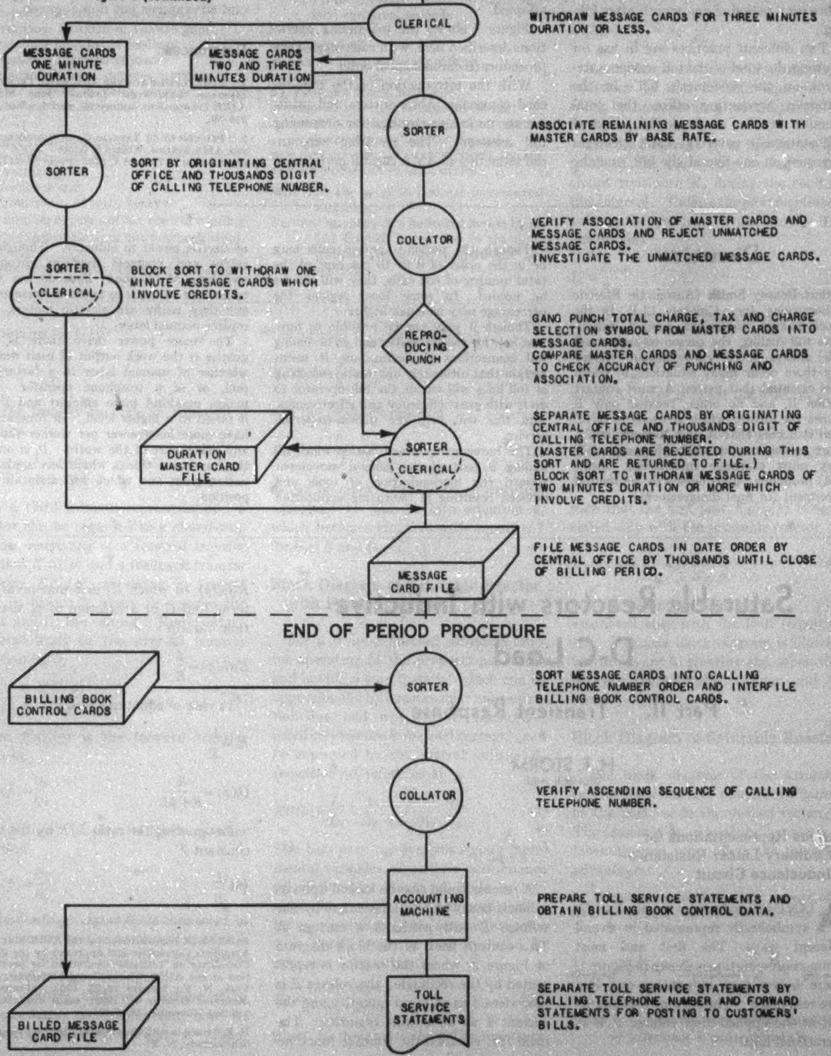
6, of the toll statement, in addition to containing the same information as the customer's portion, also shows the called area, called central office (numerical dial pulls), called telephone number, connect time, duration in minutes, and class of message.

The message cards are tabulated by billing books and the accounting machines automatically stop when billing book control cards are encountered. At this point the registers are emptied and the following data are printed on the next blank form:

Calling central office code
Billing book number
Number of messages
Number of toll users
Total toll charges
Total tax
Total charges including tax

The toll service statements are then

Figure 7 (continued)


CLERICAL
WITHDRAW MESSAGE CARDS FOR THREE MINUTES DURATION OR LESS.
MESSAGE CARDS ONE MINUTE DURATION
MESSAGE CARDS TWO AND THREE MINUTES DURATION
SORTER
ASSOCIATE REMAINING MESSAGE CARDS WITH MASTER CARDS BY BASE RATE.
SORTER
SORT BY ORIGINATING CENTRAL OFFICE AND THOUSANDS DIGIT OF CALLING TELEPHONE NUMBER.
COLLATOR
VERIFY ASSOCIATION OF MASTER CARDS AND MESSAGE CARDS AND REJECT UNMATCHED MESSAGE CARDS.
INVESTIGATE THE UNMATCHED MESSAGE CARDS.
SORTER
CLERICAL
BLOCK SORT TO WITHDRAW ONE MINUTE MESSAGE CARDS WHICH INVOLVE CREDITS.
REPRO-DUCING PUNCH
GANG PUNCH TOTAL CHARGE, TAX AND CHARGE SELECTION SYMBOL FROM MASTER CARDS INTO MESSAGE CARDS.
COMPARE MASTER CARDS AND MESSAGE CARDS TO CHECK ACCURACY OF PUNCHING AND ASSOCIATION.
DURATION MASTER CARD FILE
SORTER
CLERICAL
SEPARATE MESSAGE CARDS BY ORIGINATING CENTRAL OFFICE AND THOUSANDS DIGIT OF CALLING TELEPHONE NUMBER.
(MASTER CARDS ARE REJECTED DURING THIS SORT AND ARE RETURNED TO FILE.)
BLOCK SORT TO WITHDRAW MESSAGE CARDS OF TWO MINUTES DURATION OR MORE WHICH INVOLVE CREDITS.
MESSAGE CARD FILE
FILE MESSAGE CARDS IN DATE ORDER BY CENTRAL OFFICE BY THOUSANDS UNTIL CLOSE OF BILLING PERIOD.
END OF PERIOD PROCEDURE
BILLING BOOK CONTROL CARDS
SORTER
SORT MESSAGE CARDS INTO CALLING TELEPHONE NUMBER ORDER AND INTERFILE BILLING BOOK CONTROL CARDS.
COLLATOR
VERIFY ASCENDING SEQUENCE OF CALLING TELEPHONE NUMBER.
ACCOUNTING MACHINE
PREPARE TOLL SERVICE STATEMENTS AND OBTAIN BILLING BOOK CONTROL DATA.
BILLED MESSAGE CARD FILE
TOLL SERVICE STATEMENTS
SEPARATE TOLL SERVICE STATEMENTS BY CALLING TELEPHONE NUMBER AND FORWARD STATEMENTS FOR POSTING TO CUSTOMERS' BILLS.

separated by telephone number and the customer's portion is severed from the business office portion of the statement. The message cards used to prepare the toll service statements are not forwarded to the business office, since the business office portion of the toll statement contains all of the billing information shown on the cards. The message cards are retained by the punched card accounting office for a period of one month after billing.

Two different practices are in use for showing the total of the toll service statement on the customer's bill. In the Paterson accounting office, the total amount of the mechanically prepared toll statement covering AMA messages is posted on one line of the bill, and the amount of the manually prepared statement covering operator-handled messages is posted on another line of the bill. In the three other accounting offices, the two toll statements are placed in a jacket. The total of the two statements is entered on the jacket and is posted on a single line of the bill. It is believed that the practice followed by the Paterson accounting office will be generally adopted.

Figure 7 shows the toll billing operations described here, with each step of the procedure in chronological order.

With the introduction of the tape-to-card converter, AMA sorters and printers are no longer required for processing toll messages. The resulting substantial reduction in AMA capital investment and processing time are important advantages which have helped to make the mechanization program economical. Improvements in the design and functioning of AMA equipment and commercial business machines are under way and continued development can be expected. With these developments and improved operating techniques, the mechanization program should provide economies and advantages not now foreseen.

References

1. FUNDAMENTALS OF THE AUTOMATIC TELEPHONE MESSAGE ACCOUNTING SYSTEM, **John Meszar.** *AIEE Transactions*, volume 69, part I, 1950, pages 255–69.

2. PRINCIPLES OF TAPE-TO-CARD CONVERSION IN THE AMA SYSTEM, **Willard B. Groth.** *AIEE Transactions*, volume 72, part I, 1953 (*Paper 53–113*).

◆

Discussion

Arthur Bessey Smith (Automatic Electric Laboratories, Inc., Chicago, Ill.): In the epoch-making movement toward nation-wide toll dialing, the person-to-person call will always merit consideration. Every day there occur many situations in which it is essential that person A must contact person B, and no other, because only B knows the answer which A needs to have. Also it occurs that A and B must hold a short conference, in which they and only they must exchange information, ideas, and proposals, and come to a mutual agreement (or final disagreement).

Though the person-to-person calls may be in the order of 8 to 10 per cent of the total number of toll calls, they will always be needed. In some local regions the percentage may be much higher.

Though it may not be possible to forecast how far the subscriber can go in dialing toll connections to destination, it seems certain that ultimately automatic switching of toll lines will enable the toll operator to do it with great efficiency and effectiveness. And this will help the person-to-person service.

The movement toward nation-wide toll dialing is part of the general movement toward the mechanization of tools and devices, resulting in increasing application of electric power to industry. Though the relays and magnets used in automatic switching of telephone lines do not seem to amount to much in terms of horsepower, yet they really are electric motors, and replace manual labor.

The more power drive there is, the greater is the work output of each worker, whether of manual labor in a factory or mill, or of a telephone operator. This makes mankind more efficient and living is raised to a higher level. In America we have more horsepower per worker than in any other part of the world. It is one of the important things which has made our nation great and which will maintain our position.

Saturable Reactors with Inductive D-C Load

Part II. Transient Response

H. F. STORM
MEMBER AIEE

Various Representations for Ordinary Linear Resistance-Inductance Circuit

A LINEAR inductive reactor can be symbolically represented in several different ways. The first and most common presentation is shown in Figure 1, where L is a pure inductance and R a pure resistance. The current I as a function of the applied step voltage E is in operational form[1]

$$I=\frac{E}{R+pL} \qquad (1)$$

A reactor could also be looked upon as a black box which is stimulated by the voltage E and produces a current I. This concept leads to the block diagram of Figure 2, where the reactor is represented by the rectangle; the voltage E is now viewed as an input quantity and the current I as an output quantity. The ratio I/E is called the transfer function[2] $K_1G_1(p)$, of which K_1 is a time-invariant term, and $G(p)$ a function of p, that is, time variant

$$K_1G_1(p)=\frac{I}{E} \qquad (2)$$

In view of equations 1 and 2

$$K_1=\frac{1}{R} \qquad (3)$$

$$G_1(p)=\frac{R}{R+pL} \qquad (4)$$

Designating the ratio L/R by the time constant T

$$T=\frac{L}{R} \qquad (5)$$

Paper 53-20, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted October 17, 1952; made available for printing November 10, 1952.

H. F. STORM is with the General Electric Company, Schenectady, N. Y.

Nomenclature

$\mathfrak{a}$ = per unit ampere-turns
E = supply voltage, average
e = supply voltage, instantaneous
E_C = control voltage, average
E_F = feedback voltage, average
E_O = (nominal) output voltage, average
e_O = (nominal) output voltage, instantaneous
E_m = supply voltage, maximum
f = supply frequency, cycles per second
$F(\mathfrak{a})$ = free-wheeling function (equation 14 of part I[7])
G = response function (time-variant)
G_E = response function of voltage gain
G_2 = response function of feedback
G_L = response function of load
I_C = control current, amperes, average
I_G = gate current, amperes, average
I_L = load current, amperes, average
$j=\sqrt{-1}$
K = gain (time-invariant)
K_{AT} = ampere-turn gain
K_E = voltage gain
K_2 = feedback gain
L_L = inductance of load, henrys
N_C = number of turns of one control winding
N_F = number of turns of one feedback winding
N_G = number of turns of one gate winding
p = Heaviside operator, d/dt
R_C = control circuit resistance, ohms
R_L = load resistance, ohms
R_O = nominal output resistance, ohms
T = time constant, seconds
T_C = time constant of control circuit, seconds
T_L = time constant of load, seconds
T_n = time constant, double root of characteristic equation, seconds
T_1, T_2 = time constant, roots of characteristic equation, seconds
t = time, seconds
u = equation 47
ϵ = base of the natural logarithms
ζ = damping index
ϕ_A, ϕ_B = core fluxes, maxwells
ω = angular velocity, radians per second
ω_n = angular velocity of undamped natural frequency, radians per second
τ_C = time constant of control circuit, cycles
τ_L = time constant of load, cycles

Bars over the symbols indicate incremental variables.

Per-unit quantity $\mathfrak{a}$ is indicated in text by an Old English letter, and in the illustrations by a bar underneath the symbol.

one obtains for $G_1(p)$

$$G_1(p)=\frac{1}{1+pT} \tag{6}$$

and the transfer function of equation 2 becomes

$$K_1G_1(p)=\frac{1}{R}\,\frac{1}{1+pT} \tag{7}$$

In a third form of representation, a reactor can be regarded as a closed-loop system, consisting of a forward transfer function $K_1G_1(p)$ and a feedback transfer function $K_2G_2(p)$, as shown in general form in Figure 3, where Θ_i signifies the input and Θ_o the output. Rudimentary algebra[3] leads to the over-all transfer function Θ_o/Θ_i

$$\frac{\Theta_o}{\Theta_i}=\frac{K_1G_1(p)}{1+K_1G_1(p)K_2G_2(p)} \tag{8}$$

where $K_1G_1(p)$ is the forward transfer function

$$K_1G_1(p)=\frac{\Theta_o}{\Theta_e} \tag{9}$$

and $K_2G_2(p)$ is the feedback transfer function

$$K_2G_2(p)=\frac{\Theta_F}{\Theta_o} \tag{10}$$

The feedback signal Θ_F is connected in such manner to the summation point 1 so as to subtract from the input signal Θ_i, as indicated by the signs.

For reasons that will become more obvious later, the reactor can now be represented as shown in Figure 4. Then, the forward transfer function is

$$K_1G_1(p)=\frac{I}{E_e}=\frac{1}{pL} \tag{11}$$

and the feedback function

$$K_2G_2(p)=\frac{E_F}{I}=R \tag{12}$$

as can be quickly proved by substituting equations 11 and 12 into equation 8, which becomes identical with equation 1 for $\Theta_i=E$ and $\Theta_o=I$.

Block Diagram of Saturable Reactor with Purely Resistive Load

For a series-connected saturable reactor operating in the proportional region and having a core material which can be approximated by a rectangular magnetization line, and with further restrictions listed in reference 4, the load current $\bar{I}_L$ can be expressed by the control voltage $\bar{E}_C$ (equation 40, reference 4)

$$KG(p)=\frac{\bar{I}_L}{\bar{E}_C}=\frac{1}{R_C}\,\frac{1}{1+pT_C}\,\frac{N_C}{N_G} \tag{13}$$

The bars over the symbols signify incremental variables, as used for performance calculations by means of Bode or Nyquist diagrams.[5] The numerical value for time constant T_C of a saturable reactor with purely resistive load is listed in reference 4 (equation 35)

$$T_C=\frac{1}{4f}\,\frac{R_O}{R_C}\,\frac{N_C^2}{N_G^2} \tag{14}$$

where R_o designates the entire resistance of the saturable reactor circuit between the a-c terminals.

In analogy to Figures 3 and 4, the saturable-reactor circuit can now be represented by the block diagram of Figure 5, where

$$K_1G_1(p)=4f\,\frac{1}{R_O}\,\frac{N_G}{N_C}\,\frac{1}{p} \tag{15}$$

and

$$K_2G_2(p)=\frac{E_F}{\bar{I}_L}=R_C\,\frac{N_G}{N_C} \tag{16}$$

Again, the proof consists in substituting equations 15 and 16 into equation 8, which becomes identical with equation 13 for $\Theta_i=\bar{E}_C$ and $\Theta_o=\bar{I}_L$.

The load current $\bar{I}_L$ is the immediate result of the line voltage acting on the circuit resistance R_O during the conduction interval. Calling the aforementioned voltage the nominal output voltage $\bar{E}_O$, one obtains

$$I_L=\frac{E_O}{R_O} \tag{17}$$

or in the form of a transfer function

$$K_LG_L(p)=\frac{I_L}{E_O}=\frac{1}{R_O} \tag{18}$$

where $K_LG_L(p)$ is called the load transfer function.

One can now split the transfer function $K_1G_1(p)$ of equation 15 into two parts

$$K_1G_1(p)=K_EG_E(p)\,K_LG_L(p) \tag{19}$$

such that the first part $K_EG_E(p)$ is associated only with the saturable reactor

$$K_EG_E(p)=\frac{E_O}{E_e}=4f\,\frac{N_G}{N_C}\,\frac{1}{p} \tag{20}$$

and the second part, namely $K_LG_L(p)$ is associated only with the load (equation 18). The new block diagram is illustrated in Figure 6, showing the separation between the saturable reactor and its load.

Block Diagram of Saturable Reactor

The block diagram of the saturable reactor itself emerges now from Figure 6 as the part inside the dashed rectangle. The presentation of a saturable reactor in block-diagram form offers the following advantages:

1. When a saturable reactor is part of a larger feedback control system, such as a

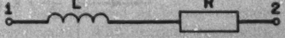


Figure 1. Linear inductive reactor, consisting of inductance L and resistance R

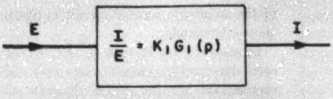


Figure 2 (above). Linear inductive reactor, represented by block with transfer function $K_1G_1(p)$

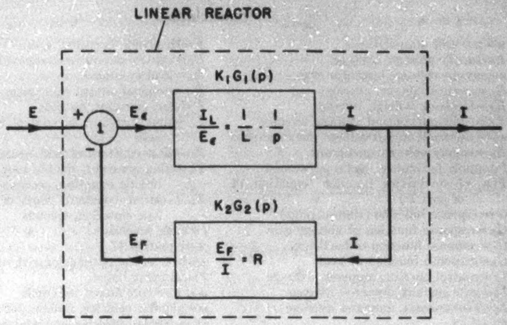


Figure 4 (right). Linear inductive reactor of Figure 1, represented by a forward function $K_1G_1(p)$, a feedback function $K_2G_2(p)$, and a summing point 1

voltage regulator, the representation of a saturable reactor in block-diagram form permits its immediate integration into the block diagram of the over-all system.

2. The representation of the saturable reactor in block-diagram form offers immediate capitalization on the techniques already developed in conjunction with feedback control systems when analyzing the transient behavior of a saturable reactor with inductive d-c load.

3. The servo block diagram offers a better insight into the mechanics of operation of saturable reactors.

FORWARD (VOLTAGE) TRANSFER FUNCTION $K_EG_E(p)$

By letting $R_C=0$, the feedback function $K_2G_2(p)$ (equation 16) vanishes, and the load current becomes a function of the forward function only. At the instant of application of the incremental control step voltage $\bar{E}_C$, and this instant can be chosen at random, the core fluxes φ_A, φ_B begin to rise, and rise linearly with time

$$\bar{\varphi}_A+\bar{\varphi}_B=\bar{E}_C\frac{10^8}{N_C}t \tag{21}$$

From equations 24, 10, and 11 of reference 4, and equations 17 and 21 of this paper, one arrives with certain reservations at

$$\bar{E}_O=4f\bar{E}_C\frac{N_G}{N_C}t \tag{22}$$

The nominal output voltage E_O is defined as the average value of the instantaneous, nominal output voltage e_O, when averaged over half a cycle. In view of this definition, the time t in equation 22 cannot mean any time. The most elementary stipulation would be to make t a multiple of π/ω which is half a cycle. It can be shown, however, that more general values for the time t can be obtained by taking the time differences between the application of the step voltage $\bar{E}_C$ and the instants when the a-c line voltage reaches crest values. $\bar{E}_O$ can now be determined from equation 22 with values for t as specified in the previous sentence, and when plotted as a function of t on linear graph paper, the points representing $\bar{E}_O$ are located on a straight line; moreover, if this line is extended toward the start until it intersects the time axis, (abscissa) the intersection occurs at exactly the point of application of the step function $\bar{E}_C$. In other words, when viewing the train of $\bar{E}_O$ values, it looks as if this train started exactly with the application of $\bar{E}_C$. This means that as far as the train of $\bar{E}_o$ values as a whole is concerned, there is no transportation lag between the inception of the step function $\bar{E}_C$ and the start of that train.

With these qualifications in mind, one may state that the output voltage $\bar{E}_O$ rises linearly with time (equation 22) regardless of the instant at which the step function is applied.

In view of

$$\frac{1}{p}=t \tag{23}$$

the forward transfer function $K_EG_E(p)$ (equation 20) becomes identical with equation 22, and all that has been stated in connection with equation 22 is valid for equation 20. Since voltage amplification is involved, $K_EG_E(p)$ will be called the voltage transfer function, and on account of the form of $G_E(p)$

$$G_E(p)=\frac{1}{p} \tag{24}$$

the transfer function $K_EG_E(p)$ is called an integrating function.

FEEDBACK TRANSFER FUNCTION $K_2G_2(p)$

During the conduction interval, one of the two saturable reactors (Figure 3, reference 4) is unsaturated and the sum of ampere-turns is zero (equation 9(*A*), reference 4).

$$I_C=I_G\frac{N_G}{N_C} \tag{25}$$

With a series-connected saturable reactor and resistive load

$$I_G\equiv I_L \tag{26}$$

and therefore the load current I_L, substituting the gate current I_G, may be shown in Figure 6 and in equation 16 as directly operating on the feedback function R_CN_G/N_C.

As a result of the transformation (equation 25) a current $\bar{I}_C$ flows in the control winding, requiring a voltage $\bar{E}_F$

$$\bar{E}_F=\bar{I}_CR_C \tag{27}$$

This voltage $\bar{E}_F$ substracts from $\bar{E}_C$ (Figure 6), leaving only $\bar{E}_\epsilon$ to the voltage transfer function. Since the voltage $\bar{E}_F$ subtracts from $\bar{E}_C$, the feedback is called negative. The saturable reactor can be viewed as a device which tries to con-

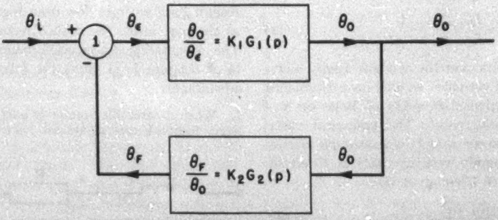


Figure 3. Block diagram, showing forward function $K_1G_1(p)$, feedback function $K_2G_2(p)$, and summing point 1

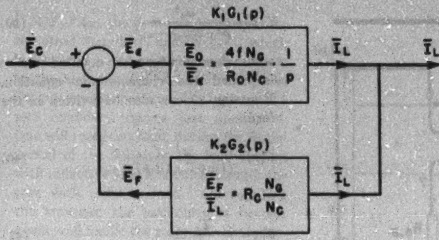


Figure 5. Block diagram of the saturable reactor with resistive load

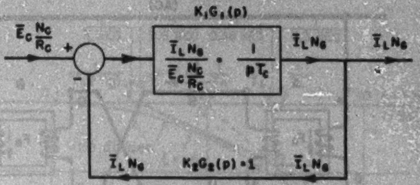


Figure 7. Block diagram of the saturable reactor with resistive load as ampere-turn amplifier

tinuously increase its output voltage $\bar{E}_O$, but which is held in check by the negative feedback of voltage $\bar{E}_F$.

For the sake of completeness, it is pointed out that equation 13 can also be written as follows

$$K_{AT}G(p)=\frac{I_L N_G}{E_C\dfrac{N_C}{R_C}}=\frac{1}{1+pT_C} \tag{28}$$

leading to the block diagram of Figure 7 for the entire circuit. Both output and input are now expressed in ampere-turns and therefore the saturable-reactor circuit can be viewed as an ampere-turn amplifier. The steady-state ampere-turn gain is

$$K_{AT}=1 \tag{29}$$

and the transient response is

$$G(p)=\frac{1}{1+pT_C} \tag{30}$$

with T from equation 14.

The feedback function $K_2G_2(p)$ becomes

$$K_2G_2(p)=1 \tag{31}$$

This feedback function of unity subtracts at the summing point from the input, and one concludes that the saturable reactor with resistive load has an ampere-turn feedback gain of minus one.

General Transfer Function of Saturable Reactor with Inductive D-C Load

The circuit is shown in Figure 8 and is described in detail in reference 7. Its block diagram is illustrated in Figure 9, which is similar to Figure 6 except as follows:

1. The load transfer of an inductive load is (equation 7)

$$K_LG_L(p)=\frac{I_L}{E_O}=\frac{1}{R_O}\,\frac{1}{1+pT_L} \tag{32}$$

where R_O signifies the entire circuit resistance when lumped into the load resistance (equation 28, reference 7). The load time constant T_L is

$$T_L=\frac{L_L}{R_O} \tag{33}$$

2. The gate current $\bar{I}_G$ no longer equals $\bar{I}_L$ as shown in equation 26. Instead

$$\frac{I_G}{I_L}=\frac{1}{F'(\alpha)} \tag{34}$$

as follows from equations 14 through 17 of reference 7. As a result, the feedback transfer function $K_2G_2(p)$ now becomes different from equation 16.

$$K_2G_2(p)=\frac{E_F}{I_L}=\frac{E_F}{I_G}\,\frac{I_G}{I_L}=R_C\,\frac{N_G}{N_C}\,\frac{1}{F'(\alpha)} \tag{35}$$

where $F'(\alpha)$ is the derivative of the freewheeling function $F(\alpha)$ with respect to, α (equations 12, 13, and 14 and 38, reference 7).

With all component transfer functions established (see Figure 9), the total transfer function can be obtained from equation 8

$$\frac{I_L}{E_C}=\frac{F'(\alpha)}{R_C}\,\frac{N_C}{N_G}\cdot\left[\frac{1}{p^2T_LT_CF'(\alpha)+pT_CF'(\alpha)+1}\right] \tag{36}$$

where T_C is the time constant of the control circuit of the saturable reactor in the absence of inductance in the load circuit (equation 14).

As in the foregoing, the time-invariant part K is the steady-state gain. The response function $G(p)$ is in a form which has been solved in the past in connection with feedback control problems.[8–11] A formidable amount of time can now be saved by the direct application of the results of the analysis of feedback control systems to the saturable reactor with inductive d-c load.

The response function $G(p)$ has variable coefficients, a fact which makes a rigorous solution very difficult. However, for the evaluation of stability in feedback control problems, the response needs to be determined only in the immediate vicinity of a steady-state operating point. Then, for the small excursions of the incremental variables, to which this analysis is restricted, the response function can be considered as having constant coefficients. With constant coefficients, the quadratic response function $G(p)$ represents a second-order differential equation similar to the equation for the

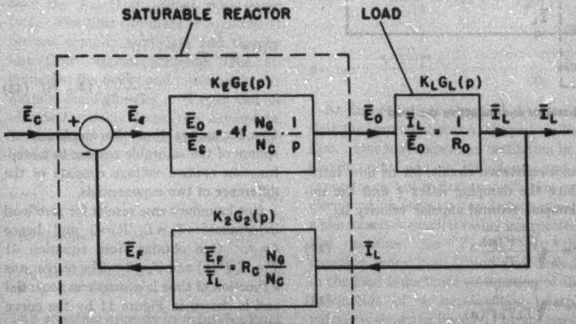


Figure 6. Block diagram of the saturable reactor and its interconnection to a separately shown resistive load

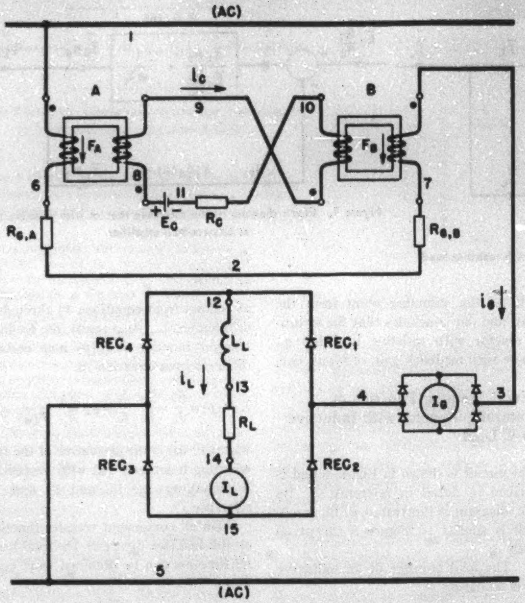

(AC)
1
A
I_C
9
10
B
F_A
F_B
6
8
7
11
E_C
R_C
$R_{G,A}$
$R_{G,B}$
2
12
I_G
REC_4
L_L
REC_1
I_L
13
4
3
I_G
R_L
14
REC_3
REC_2
I_L
15
5
(AC)

Figure 8. Schematic diagram of series-connected saturable reactor with inductive d-c load

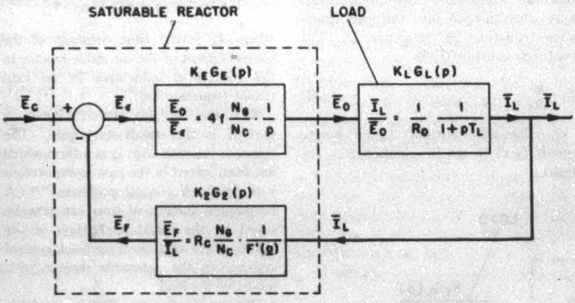

SATURABLE REACTOR
LOAD
$K_E G_E(p)$
$K_L G_L(p)$
$\bar{E}_C$
+
-
$\bar{E}_\epsilon$
$\frac{\bar{E}_O}{\bar{E}_\epsilon} = 4f\frac{N_G}{N_C}\cdot\frac{1}{p}$
$\bar{E}_O$
$\frac{\bar{I}_L}{\bar{E}_O} = \frac{1}{R_O}\cdot\frac{1}{1+pT_L}$
$\bar{I}_L$
$\bar{I}_L$
$K_2 G_2(p)$
$\bar{E}_F$
$\frac{\bar{E}_F}{\bar{I}_L} = R_C\frac{N_G}{N_C}\cdot\frac{1}{F'(\mathbf{a})}$
$\bar{I}_L$

Figure 9. Block diagram of saturable reactor and inductive d-c load

charge of a capacitor as a function of the terminal voltage in a linear series network consisting of capacitance, inductance, and resistance. The response of such a system to a step-function disturbance of the terminal voltage can be classified into overdamped response, underdamped (oscillatory) response, and critically damped response for the boundary between the two aforementioned cases.

In analogy to the capacitance-inductance-resistance circuit, let us now introduce the damping index ζ and the undamped, natural angular velocity ω_n

$$\zeta=\frac{1}{2}\sqrt{\frac{T_C F'(\mathbf{a})}{T_L}} \tag{37}$$

$$\omega_n=\sqrt{\frac{1}{T_L T_C F'(\mathbf{a})}} \tag{38}$$

With these substitutions, the response function of equation 36 becomes

$$G(p)=\frac{\omega_n^2}{p^2+2\zeta\omega_n p+\omega_n^2} \tag{39}$$

whose denominator when equated to zero is called the characteristic equation. Equation 39 can also be written in the form

$$G(p)=\frac{1}{(1+pT_1)(1+pT_2)} \tag{40}$$

where

$$T_{1,2}=\frac{T_n}{\zeta\mp\sqrt{\zeta^2-1}} \tag{41}$$

and

$$T_n=\frac{1}{\omega_n} \tag{42}$$

OVERDAMPED RESPONSE ($\zeta>1$)

With $\zeta>1$, the time constants $T_{1,2}$ (equation 41) are real and different. The response function $G(j\omega)$ can be easily plotted in the attenuation-frequency plane.[10] The two time constants T_1 and T_2 are different from $T_C F'(\mathbf{a})$ and T_L and lie between $T_C F'(\mathbf{a})$ and T_L. A practical case is shown in Figure 10 where $T_C F'(\mathbf{a}) = 3$ seconds and $T_L = 1/3$ second. The damping ratio $\zeta = 1.5$ (equation 37). The two time constants follow from equation 41: $T_1 = 2.62$ seconds, $T_2 = 0.38$ second. The asymptotes of the attenuation-frequency diagram can now be plotted by the well-known procedure,[9,10] resulting in the three straight lines *OABC*. For geometric reasons, the extension of *CB* intersects the zero-decibel axis at point *D*, which is located at $\omega=\omega_n$. With the asymptotes established, the attenuation-frequency characteristic and the phase-angle–frequency characteristic can be quickly obtained by simple standard methods.

The operational form of equation 40 can be converted into an amplitude-time function,[12] resulting in

$$G(t)=\frac{T_1}{T_1-T_2}\left(1-\epsilon^{-\frac{t}{T_1}}\right)-\frac{T_2}{T_1-T_2}\left(1-\epsilon^{-\frac{t}{T_2}}\right) \tag{43}$$

One recognizes that the amplitude response of the saturable reactor to a step-function control voltage consists of the difference of two exponentials.

One boundary case results for zero load inductance: $T_L = L_L/R_O = 0$ and hence $\zeta=\infty$. One obtains from equation 41 $T_1 = T_C F'(\mathbf{a})$ and $T_2 = 0$. The response as a function of time is a simple exponential and is shown in Figure 11 by the curve marked $\zeta=\infty$.

As the load inductance L_L is increased, the damping index declines. For $T_L=$

$(1/4)T_CF'(\mathbf{a})$ the damping index reaches the other boundary, namely $\zeta=1$, which is discussed in the next section and shown by the curve marked $\zeta=1$. The response for any overdamped case lies between the two boundary curves just mentioned. One will recognize that during the initial period of the transient the load current with inductive load is smaller than for the case with resistive load. Later during the transient the two currents become equal, and finally the current in the inductive case stays ahead of the current in the resistive case. The latter conclusion is rather arresting, and this seeming paradox of a more rapid current build-up in the inductive case becomes more conspicuous yet for $\zeta>1$.

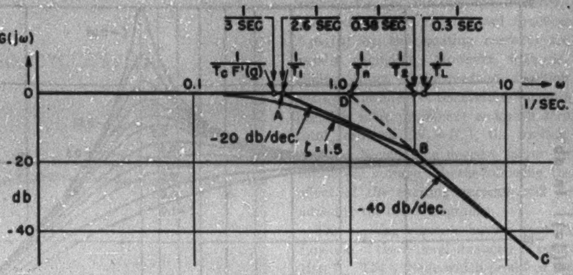


Figure 10. Attenuation frequency response of load current for overdamped system

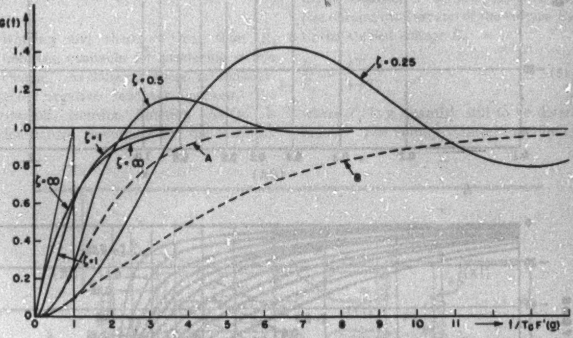


Figure 11. Amplitude-time response of load current for various degrees of damping

CRITICALLY DAMPED RESPONSE ($\zeta=1$)

By increasing the load inductance L_L until $T_L=(1/4)T_CF'(\mathbf{a})$, the damping index $\zeta=1$ and one now obtains coinciding values for the two time constants (equation 41) corresponding to a double root of the characteristic equation. The response function $G(p)$ degenerates into

$$G(p)=\frac{1}{(1+pT_n)^2} \tag{44}$$

The amplitude time response[12] corresponding to equation 44 is

$$G(t)=1-\epsilon^{-\frac{t}{T_n}}\left(1+\frac{t}{T_n}\right) \tag{45}$$

and is illustrated in Figure 11. The critically damped case is characterized by the load current reaching the vicinity of its asymptote $G(t)=1$ most quickly without causing an overshoot.

UNDERDAMPED OSCILLATORY RESPONSE ($\zeta<1$)

By further increasing the load inductance L_L, $T_L>(1/4)T_CF'[\mathbf{a}]$, and hence $\zeta<1$. The time constants of equation 41 become complex conjugate and thereby lose their original meaning. The attenuation-frequency and phase-angle–frequency characteristics no longer can be obtained quickly from the asymptote. By replacing p by $j\omega$ in equation 39 and by substituting

$$u=\frac{\omega}{\omega_n} \tag{46}$$

one obtains

$$G(ju)=\frac{1}{-u^2+2\zeta ju+1} \tag{47}$$

This equation appears in many feedback control problems and its numerical evaluation has been repeatedly published;[11] it is again shown in Figure 12 for various damping indices.

By transforming the attenuation-frequency response into the amplitude-time response,[12] one obtains equations 48 and 49

$$G(t)=1-\frac{\epsilon^{-\zeta\omega_nt}}{\sqrt{1-\zeta^2}}\sin\left(\omega_n\sqrt{1-\zeta^2}\,t+\phi\right) \tag{48}$$

$$\phi=\tan^{-1}\frac{\sqrt{1-\zeta^2}}{\zeta} \tag{49}$$

An inspection of equation 48 reveals that the load current is a damped oscillation. One may expect an oscillation in a circuit containing an inductance and a capacitance, but to find an oscillation in a circuit which contains two inductances is a phenomenon which is rather unexpected. For instance, for $T_L=T_CF'(\mathbf{a})$, the damping index $\zeta=0.5$, and the current in the load inductance as function of the application of a step-function control voltage is shown in Figure 11 by the curve marked $\zeta = 0.5$. For a larger T_L, namely $T_L = 4\ T_CF'(\mathbf{a})$, the curve marked $\zeta=0.25$ results. The load current overshoots and undershoots the asymptote $G(t)=1$ and gradually approaches it in the form of a damped oscillation.

Not only is the oscillatory character of the load current quite unexpected but the rate of current rise is another surprising feature. Suppose one energizes the field of an exciter from the armature of a pilot exciter, each machine having a transfer function of the general form $1/(1+pT)$. The response of the exciter field current as a function of a step voltage applied to the field of the pilot exciter is then given by equation 43, where T_1 is the field time constant of the pilot exciter and T_2 is the exciter field time constant. For $T_2=T_1$, the exciter field current will respond according to the dashed curve A in Figure 11 with t/T_1 as abscissa. However, if the same exciter field is energized from a saturable reactor such that $T_CF'(\mathbf{a})=T_1$, the curve marked $\zeta=0.5$ results, as previously discussed. For $T_2=4T_1$ the exciter field current responds according to curve B when energized from the aforementioned pilot exciter. However, when

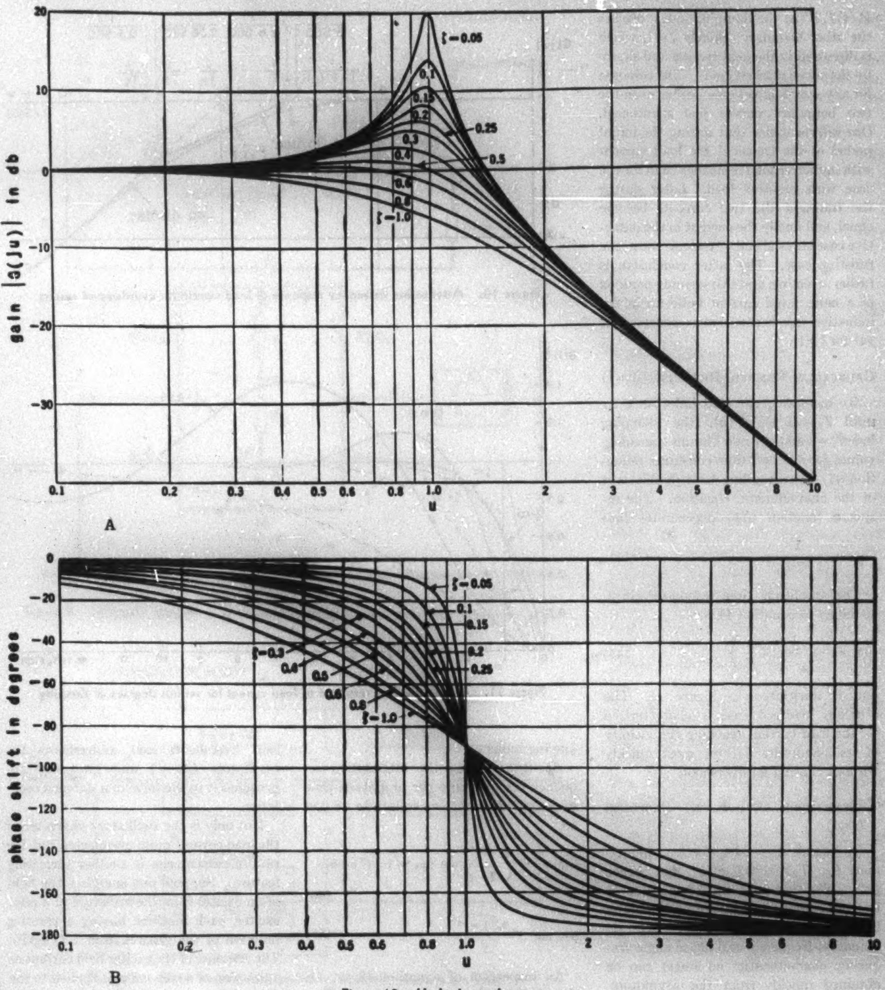


Figure 12. Underdamped system

A. Attenuation-frequency response of load current

B. Phase-angle–frequency response of load current

energized by the aforementioned saturable reactor, the exciter field current responds as shown by the curve marked $\zeta = 0.25$. One concludes that, with equal time constants, a saturable reactor taking the place of the pilot exciter gives a much quicker response of exciter field current, and hence exciter armature voltage.

Figure 13 illustrates the load current i_L through the load inductance as the result of a suddenly increased control voltage E_C of the saturable reactor, energizing the load inductance (Figure 9). The upper trace in the oscillogram is a time wave obtained from the 60-cycle supply voltage. The measured value for the load time constant τ_L equals 9 cycles, the control time constant τ_C equals 2.5 cycles and the average value of $F'(a)$ equals 1.6, resulting in a damping ratio of $\zeta = 0.33$. The ratio of the maximum amplitude of the load current with respect to the final steady-state value is measured from the oscillogram as 1.35. The calculation of the maximum overswing from equations 48 yields 1.33, which is in close agreement with the measured value.

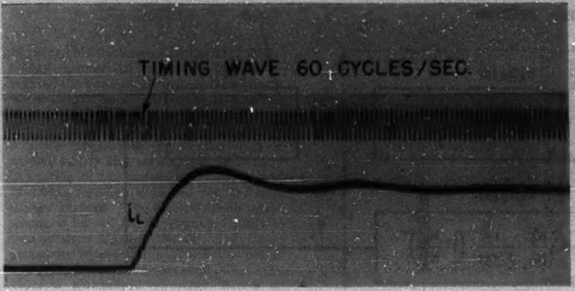


Figure 13. Load current i_L in inductive d-c load, as result of suddenly increased control voltage E_C of saturable reactor

Block Diagram of Saturable Reactor with External Feedback (Also Amplistats)

As shown in Figure 14, feedback windings (*Z1*, *Z2*) and a feedback rectifier bridge have been added to Figure 8. The rectified gate current flowing in the feedback windings induces opposite ampere-turns in the control windings, with the result that the control current becomes reduced by the factor $1-(N_F/N_G)$ (equation 39, reference 13). Then the feedback transfer function $K_2G_2(p)$ (equation 35) becomes reduced by the same factor (Figure 15)

$$K_2G_2(p)=R_C\frac{N_G}{N_C}\,\frac{1}{F'(n)}\left(1-\frac{N_F}{N_G}\right) \qquad (50)$$

For $N_F=N_G$, the external positive feedback completely nullifies the internal negative feedback, resulting in an amplifier without feedback; the feedback link as represented by $K_2G_2(p)$ in Figure 15 vanishes, and the forward function $K_BG_B(p)$ becomes the total transfer function of the magnetic amplifier. Instead of applying external positive feedback for the cancellation of the inherent negative feedback, the latter can be eliminated more directly by the interposition of rectifiers. These rectifiers block the flow of gate current in that direction which would cause the negative feedback; such circuits are called amplistats,[14] self-saturated,[18] self-feedback[19] amplifiers.

The forward function $K_BG_B(p)$ (equation 20) has been derived under the assumption of the theoretical rectangular flux-density–field-intensity line which, by its very nature, will not engender exciting currents. With actual core materials, however, exciting currents are present. These exciting currents cause voltage drops on the resistance R_C of the control winding and therefore less than $\bar{E}_e$ becomes available for producing a flux change. As long as there is considerable negative feedback present, the over-all transfer function $KG(p)$ is determined essentially by the feedback function $K_2G_2(p)$, and therefore the deviations inflicted to the forward function $K_BG_B(p)$ by the exciting currents are usually negligible. However, with declining values of the feedback function $K_2G_2(p)$, the effect of the exciting currents becomes more prominent and with zero total feedback ($N_F=N_G$), the exciting currents become the decisive factor for the output voltage. In the following, the effect of the exciting currents will be added to the block diagram.

The exciting currents have a transfer function of the general character of equation 7. The voltage which produces the flux change is the voltage on the inductance E_X, which is obtained by multiplying equation 7 by pL. As a result, one obtains for the ratio of the voltage E_X to the applied voltage E_e

$$\frac{E_X}{E_e}=\frac{pT_a}{1+pT_a} \qquad (51)$$

where T_a is a quantity still to be determined. The voltage transfer function of

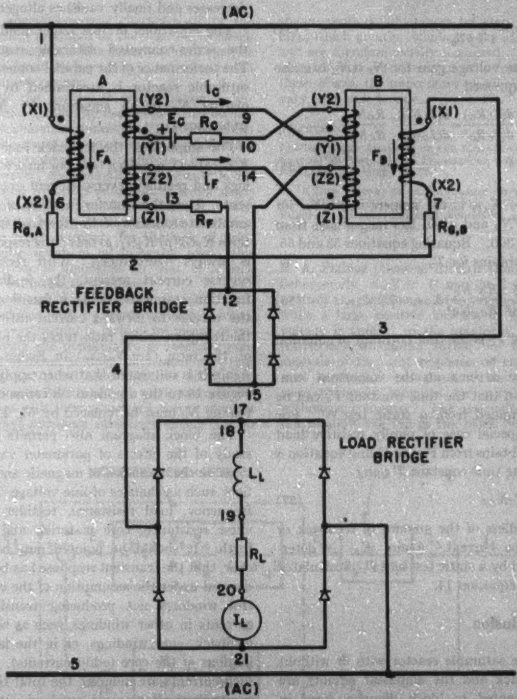


Figure 14. Schematic diagram of series-connected saturable reactor with external gate current feedback, and with inductive d-c load

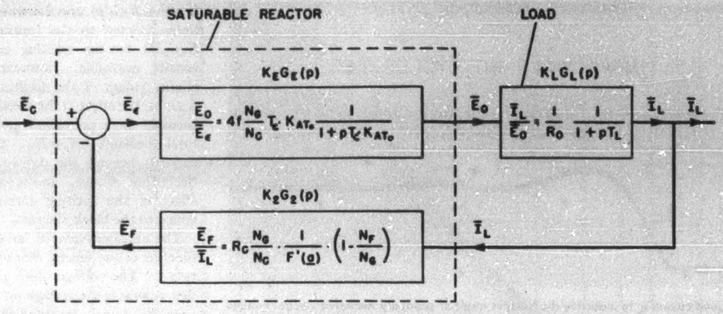

SATURABLE REACTOR
LOAD
$K_E G_E(p)$
$K_L G_L(p)$
$\bar{E}_C$
+
−
$\bar{E}_\epsilon$
$\frac{\bar{E}_O}{\bar{E}_\epsilon} = 4f \frac{N_G}{N_C} T_C \cdot K_{AT_0} \frac{1}{1 + pT_C K_{AT_0}}$
$\bar{E}_O$
$\frac{\bar{I}_L}{\bar{E}_O} = \frac{1}{R_O} \frac{1}{1 + pT_L}$
$\bar{I}_L$
$\bar{I}_L$
$K_2 G_2(p)$
$\bar{E}_F$
$\frac{\bar{E}_F}{\bar{I}_L} = R_G \frac{N_G}{N_G} \cdot \frac{1}{F'(\underline{a})} \cdot \left(1 - \frac{N_F}{N_G}\right)$
$\bar{I}_L$

Figure 15. Block diagram of series-connected saturable reactor with external gate current feedback, and with inductive d-c load. $K_2G_2(p)=0$ for amplistats with zero reverse-rectifier current

equation 20 becomes

$$\frac{E_O}{E_\epsilon} = 4f \frac{N_G}{N_C} \frac{1}{p} \frac{pT_a}{1+pT_a} \tag{52}$$

$$K_E = 4f \frac{N_G}{N_C} T_a \tag{53}$$

$$G_E(p) = \frac{1}{1+pT_a} \tag{54}$$

The voltage gain for $N_F = N_G$ can also be expressed

$$K_E = \frac{d(I_L R_O)}{d(I_C R_C)} = \frac{d(I_L N_G)}{d(I_C N_C)} \frac{R_O N_C}{R_C N_G} = K_{AT_0} \frac{R_O N_C}{R_C N_G} \tag{55}$$

where K_{AT_0} is the ampere-turn gain for $N_F = N_G$ and often has magnitudes from 30 to 300. Equating equations 53 and 55, one obtains for T_a

$$T_a = \frac{1}{4f} \frac{R_O}{R_C} \frac{N_C^2}{N_G^2} K_{AT_0} = T_C K_{AT_0} \tag{56}$$

where T_C is the time constant of equation 14.

One arrives at the important conclusion that the time constant T_a can be determined from a static test.[16,17] For the special case of purely resistive load one obtains from Figure 15 and equation 8 for the time constant T

$$T = T_C K_{AT} \tag{57}$$

regardless of the amount of feedback or reverse current. Again, K_{AT} is determined by a static test and T_C is calculated from equation 14.

Conclusion

The saturable reactor with or without feedback and the amplistat circuits are represented in block diagram form in Figure 15. The load is either a purely resistive or an inductive d-c load. The transient response is obtained from the block diagram by a simple manipulation, equation 8. With inductive d-c load, the response of the load current may be oscillatory, Figure 13; as positive feedback is increased, the oscillatory character decreases and finally vanishes altogether.

The equations of this paper hold for the series-connected saturable reactor. The performance of the parallel-connected saturable reactor is ascertained by replacing N_G in all equations by $N_G/2$ (reference 4).

For amplistats, the feedback function $K_2G_2(p) = 0$, provided boost-or-buck windings and rectifier reverse current are absent. Then the transfer function of the amplistate consists of the forward functions $K_E G_E(p)\ K_L G_L(p)$ only; the response is always overdamped. With rectifier reverse current present, the feedback function reappears, being a function of the reverse to forward current ratio of the rectifiers, which ratio takes the place of the term $(1 - N_F/N_G)$ in Figure 15. Again it is reiterated that when applying Figure 15 to the amplistat all terms containing N_G must be replaced by $N_G/2$.

The block diagram also permits the study of the effects of parameter variations on the operation of magnetic amplifiers, such as changes of line voltage, line frequency, load resistance, rectifier reverse resistance, core material, and so forth. It should be pointed out, however, that the transient response has been derived under the assumption of the control windings not producing transient currents in other windings, such as bias, feedback, gate windings, or in the laminations of the core (eddy currents). If such currents are induced, the total time constant of all currents involved enters into the calculation, and a considerable complication of the analysis results.

References

1. Heaviside's Operational Calculus (book), **E. J. Berg.** McGraw-Hill Book Company, Inc., New York, N. Y., 1936, page 3.

2. Principles of Servomechanisms (book), **G. S. Brown, D. P. Campbell.** John Wiley and Sons, Inc., New York, N. Y., 1948, page 85.

3. *Ibid.*, page 143.

4. Transient Response of Saturable Reactors with Resistive Load, **H. F. Storm.** *AIEE Transactions*, volume 70, part I, 1951, pages 95–102.

5. Servomechanisms and Regulating System Design (book), **H. Chestnut, R. W. Mayer.** John Wiley and Sons, Inc., New York, N. Y., volume 1, 1951, page 291.

6. **Berg,** *op. cit.*, page 5.

7. Saturable Reactors with Inductive D-C Load. Part I—Steady-State Operation, **H. F. Sterm.** *AIEE Transactions*, volume 71, part I, November 1952, pages 335–43.

8. Servomechanism Fundamentals (book), **H. Lauer,** *et al.*, McGraw-Hill Book Company, Inc., New York, N. Y., 1947, page 63.

9. Theory of Servomechanisms (book), **H. M. James** *et al.*, McGraw-Hill Book Company, Inc., New York, N. Y., 1947, pages 143, 175.

10. **Brown, Campbell,** *op cit.*, page 236.

11. **Chestnut, Mayer,** *op. cit.*, page 310.

12. **Berg,** *op. cit.*, page 164.

13. Series Connected Saturable Reactor with Control Source of Comparatively Low Impedance, **H. F. Storm.** *AIEE Transactions*, volume 69, part II, 1950, pages 756–65.

14. The Amplistat and its Application, **H. M. Ogle.** *General Electric Review* (Schenectady, N. Y.), volume 53, February, August, October 1950.

15. The Behaviour of a Series-Transductor Magnetic Amplifier With Directly-Connected or Rectified-Fed Loads, **A. G. Milnes.** *Proceedings*, Institution of Electrical Engineers (London, England), volume 99, part II, February 1952, pages 13–23.

16. The Figure of Merit of Magnetic Amplifiers, **J. T. Carleton, W. F. Horton.** *AIEE Transactions*, volume 71, part I, September 1952, pages 239–45.

17. The Transductor Amplifier (book), **U. Krabbe.** *Lindhska Boktryckeriet* (Orebro, Sweden), 1947, page 87.

18. Self-Saturation in Magnetic Amplifiers, **M. J. Dornhoefer.** *AIEE Transactions*, volume 68, part II, 1949, pages 835–50.

19. The Effective Feedback Ratio of Magnetic Amplifiers, **L. A. Finzi, G. F. Pittman, Jr., H. L. Durand.** *AIEE Transactions*, volume 71, part I, 1952, pages 157–64.

Discussion

M. Riaz (Massachusetts Institute of Technology, Cambridge, Mass.): The advantages of the block diagram have been clearly demonstrated by Dr. Storm in his analysis of the transient response of saturable reactors: not only does the technique offer a better insight into the physics of these devices but it also permits the systematic utilization of the methods of analysis and synthesis already developed for feedback control systems. The block-diagram representation has an added desirable feature in that it often discloses the similar dynamic properties of essentially different types of energy-conversion devices and thus provides a generalized approach based on analogy that is helpful in understanding the mechanics of operation of these devices. In this respect, the analogy between saturable reactors and metadynes is indeed very striking.

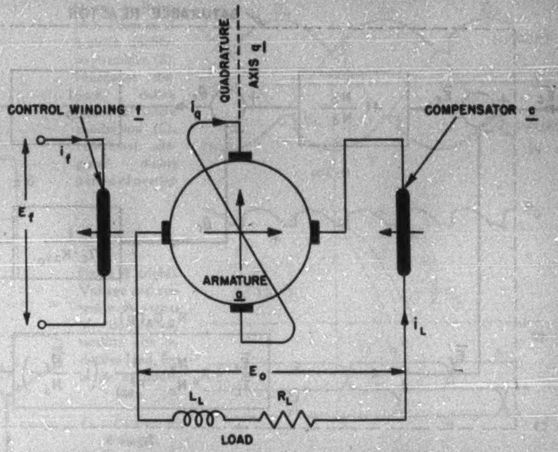


Figure 1. Schematic diagram of a simple metadyne generator supplying an inductive load

A simple metadyne generator with resistive-inductive load is shown schematically in Figure 1 of the discussion. A general transient analysis of the metadyne is given in reference 1. If we assume that the control field circuit has a very small time constant (or that it is energized from a large impedance constant-current source), and if we further neglect the mutual inductance effects, the differential equations governing the operation of the metadyne of Figure 1 of the discussion are

$$E_f = i_f R_f$$

$$E_O = K_a i_q = i_L R_O (1 + pT_L)$$

$$K_f i_f - (K_a - K_c) i_L = i_q R_q (1 + pT_q)$$

where

K = speed-voltage coefficient = knN
N = turns
n = speed of rotation
k = constant dependent upon number of conductors and permeance of the magnetic circuit

The block diagram representation of the metadyne generator shown in Figure 2 of the discussion is easily derived from these equations. It is similar to Figure 15 of the paper, which represents a saturable reactor, except for the fact that the forward gain of the saturable reactor is directly proportional to frequency, whereas the forward gain of the metadyne is proportional to the square of the speed. This squared-speed term appears in the transfer function of the metadyne because here we are essentially dealing with a 2-stage structure. It is interesting to notice the equivalence of the effects of exciting currents in both representations (the excitation field of the metadyne is produced by quadrature current i_q).

Because the same block diagram is applicable to both saturable reactor and metadyne, results obtained for one structure may also be derived for the other. Dynamically both devices are essentially similar.

The oscillatory character of the load current is a main feature of the metadyne (see Figure 13 of reference 1) whose inherent fast response makes it especially suited to its use as an exciter. This property known as field-forcing action, is a result of the constant-current characteristic of the machine produced by the negative feedback of the output. The metadyne may then be considered analogous to a series-connected saturable reactor.

As the degree of compensation is increased (feedback decreased), the oscillatory character lessens and finally vanishes in the case of perfect compensation ($N_a = N_c$) corresponding to amplidyne operation. However, the over-all amplification is increased with this reduction of scalar negative feedback. Amplidyne behavior is then dictated by its constant-voltage characteristic which accounts for the lack of field-forcing action in an inductive load; this behavior is analogous to that of a self-excited magnetic amplifier.[2]

Although block diagram techniques, often developed with electronic systems in mind, are very helpful in looking for means of improving the performance of feedback control systems, it must be remembered that they are based upon idealized conditions. Magnetic devices of both static and rotating types with their associated energy storages always introduce subsidiary effects (for instance, mutual inductances between windings) which greatly complicate the analysis and are therefore usually neglected. However, all final assessment must account for these effects, the importance of which may vary from case to case.

References

1. Transient Analysis of the Metadyne Generator, M. Riaz. *AIEE Transactions*, volume 72, February 1953, pages 52–62.

2. A New Theory of the Magnetic Amplifier, A. G. Milnes. *Proceedings*, Institution of Electrical Engineers (London, England), volume 97, part II, 1950, pages 460–74.

R. A. Phillips (General Electric Company, Schenectady, N. Y.): It may be desirable to study the performance of a control system when a large positive increase of control voltage is applied to the amplistat. The restrictions which the author put on his analysis of transient response, as enumerated in the last paragraph of the paper, were obviously intended for calculations without a computer. The restrictions do not disallow the use of the block diagram

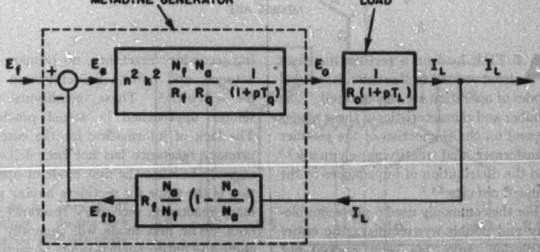

METADYNE GENERATOR
LOAD
E_f
+
−
E_e
$n^2 k^2 \frac{N_f N_a}{R_f R_q} \frac{1}{(1+pT_q)}$
E_o
$\frac{1}{R_o(1+pT_L)}$
I_L
I_L
$R_f \frac{N_a}{N_f} \left(1 - \frac{N_c}{N_a}\right)$
E_{fb}
I_L

Figure 2. Block-diagram representation of metadyne system

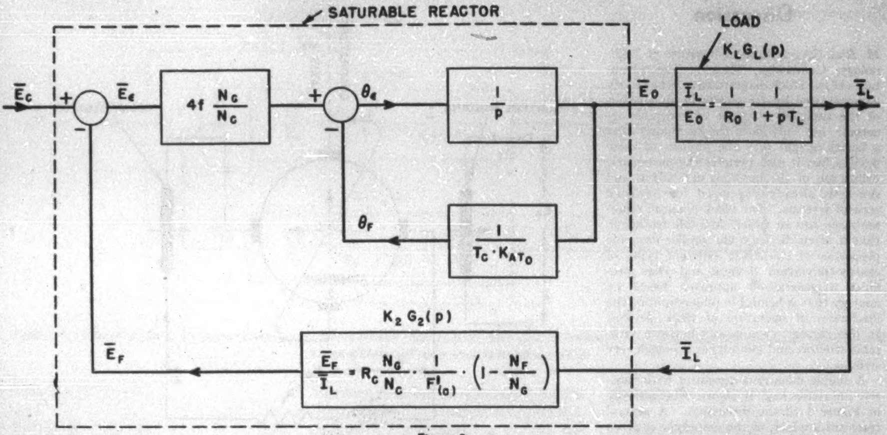

SATURABLE REACTOR
LOAD
$K_LG_L(p)$
$\bar{E}_C$
+
−
$\bar{E}_\epsilon$
$4f\frac{N_G}{N_C}$
+
−
θ_ϵ
$\frac{1}{p}$
$\bar{E}_O$
$\frac{\bar{I}_L}{E_O}=\frac{1}{R_O}\frac{1}{1+pT_L}$
$\bar{I}_L$
θ_F
$\frac{1}{T_C\cdot K_{AT_O}}$
$K_2G_2(p)$
$\bar{E}_F$
$\frac{\bar{E}_F}{\bar{I}_L}=R_C\frac{N_G}{N_C}\cdot\frac{1}{F'_{(a)}}\cdot\left(1-\frac{N_F}{N_G}\right)$
$\bar{I}_L$

Figure 3

in connection with a computer for a large positive increase of control voltage provided that the nonlinear nature of the control characteristic is included.

Suppose in a voltage regulating system a magnetic amplifier is used as the pilot exciter for the main exciter of a synchronous generator. One may wish to determine the rate of rise of the exciter for a large error signal to the regulator caused by a large power system disturbance (such as switching out a line). The saturable reactor can no longer be considered a linear device since it goes to ceiling voltage quite rapidly. The resulting nonlinear effects can be handled if an analogue computer is available or step-by-step integration is employed. The following procedure will give the solution.

Refer now to Figure 15. The forward transfer function $K_EG_E(p)$ can be written as

$$E_O=\left[4f\frac{N_G}{N_C}E_\epsilon-\frac{E_O}{T_CK_{AT_0}}\right]\frac{1}{p}$$

This is recognized as a forward transfer function and a feedback transfer function and the block diagram of Figure 15 becomes Figure 3 of the discussion.

Since K_{AT_0} is the slope of the control characteristic at any point and T_C is a constant determined from equation 14 of the paper, a curve relating θ_F to $\bar{E}_O$ can easily be drawn. A curve of $K_2G_2(p)$ of which $F'(a)$ is the only variable may be drawn as a function of $\bar{I}_L$. Nonlinear elements on the analogue computer will then permit a continuous solution. This sort of thing is done frequently to handle saturation in rotating machines. The block diagram makes the steps of analysis easier to visualize. When using a computer, the other restrictions of Dr. Storm's analysis can be obviated in the same manner.

H. F. Storm: The author wishes to thank R. A. Phillips and M. Riaz for their worthwhile discussions.

Extended Regulation Curves for 6-Phase Double-Way and Double-Wye Rectifiers

I. K. DORTORT
MEMBER AIEE

AS THE load on a rectifier increases from no load to short circuit, several modes of operation are encountered. The number and characteristics of these modes depend on the connection of the rectifier transformer and rectifying elements,[1,2] and the distribution of impedances in the coupled circuits.[3,4]

For the commonly used 6-phase double-way and double-wye rectifiers, the entire regulation curve has been worked out only for the special case of a rectifier having secondary reactance, no primary reactance, and infinite inductance in the d-c circuit.[1,2] These conditions are seldom approached in actual practice. The lack of information for the case of primary reactance has not been felt too severely because the first mode of operation is the same for rectifiers having primary reactance, secondary reactance, or both. The first mode will generally include the entire normal range of rectifier operation.

The value of the short-circuit current for the case of a rectifier with only primary reactance and no inductance in the d-c circuit is given in an AIEE report.[3] Formulas for the entire range of load on rectifiers with only primary reactance and having either infinite or zero d-c inductance were presented in a paper by the author[4] and later found to be in error in two areas. This paper presents the derivation of the regulation curve for the case of primary reactance only, and either infinite or zero inductance in the d-c circuit.

The definitions of primary and secondary reactance are important. For the purpose of this paper it is sufficient to

---

Paper 53-36, recommended by the AIEE Electronic Power Converters Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New Yook, N. Y., January 19-23, 1953. Manuscript submitted October 21, 1952; made available for printing November 17, 1952.

I. K. Dortort is with the I-T-E Circuit Breaker Company, Philadelphia, Pa.

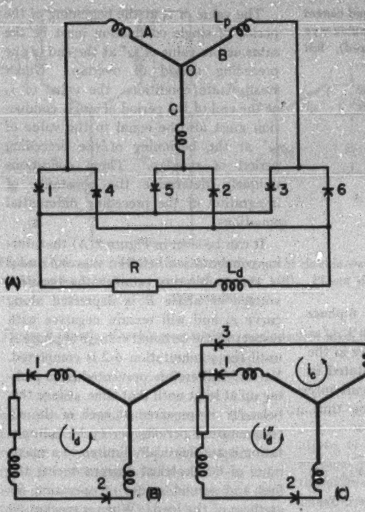


Figure 1 (left). (A). Diagram of 6-phase double-way rectifier. (B). Functional diagram during period of single conduction. (C). Functional diagram during period of overlap

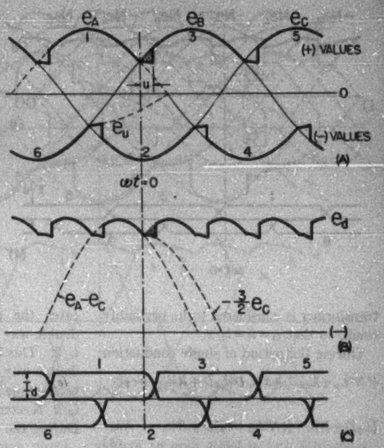


Figure 2 (right). Voltage and current wave shapes of double-way rectifier with inductive load, first mode of operation

A. Alternating voltage

B. Direct voltage

C. Valve currents

define primary reactance as reactance which produces equal voltage drops in diametrically opposite valve circuits. Secondary reactance produces a voltage drop in only one valve circuit. Generally, primary reactance will involve only alternating leakage fluxes, while secondary reactance involves only unidirectional leakage fluxes.

Definition of Symbols

As far as possible the legend of symbols has been made consistent with the nomenclature of the report on "Protection of Electronic Power Converters."[3] Subscript pu denotes per-unit quantities.

$e_{A,B,C}$ = instantaneous phase voltages
E = rms value of phase voltages
E_d = average direct voltage
ΔE_d = direct voltage drop $= E_{d0}-E_d$
E_{d0} = average direct voltage at no load
$E_{dpu}=E_d/E_{d0}$
$i_{1,2,3}$ = instantaneous valve currents
i_d = instantaneous direct current during single conduction (only one phase conducting in a given direction)
i_d'' = instantaneous direct current during overlap (two phases conducting in the same direction)
i_c = instantaneous commutating current
L_d = d-c inductance
L_p = primary inductance
I_d = average d-c load current
I_{dn} = rated d-c load current
N = turn ratio, secondary to primary
S = load factor $= X_{cpu}(I_d/I_{dn})$
F_x = reactance factor $=(I_dX_c)/E_s$
u = angle of overlap (double conduction)
α = angle of delay
γ = angle of double commutation
X_c = commutating reactance
X_{cpu} = per-unit commutating reactance $=(\sqrt{2}X_cI_{dn})/(\sqrt{3}E)$

Equality of 6-Phase Double-Wye and Double-Way

It is well known that the 6-phase double-way and double-wye circuits exhibit similar characteristics during the first mode of operations. If it can be proved for the general case that both circuits behave in identically the same manner from no load to short circuit, the derivation of the extended regulation curve for one will hold for both.

Taking first the case of the 6-phase double-way rectifier, Figure 1(A) shows the actual diagram, 1(B) shows a functional diagram for the period of single conduction, when only one phase conducts in a given direction, and 1(C) is the functional diagram during a period of overlap, u when two phases are conducting in the same direction. Figure 2 shows the characterized voltage and current wave shapes.

Using the operator $D=d/dt$, the current i_d through the d-c load during the period of single conduction, Figure 1(B) is given by the generalized equation

$$[(2L_p+L_d)D+R]i_d=e_A-e_C \qquad (1)$$

When e_B becomes equal to $e_A-L_pDi_d$, valve *3* is free to conduct current and the period of overlap u begins. The build-up of current in valve *3* and the decay of current in valve *1* can be represented by a circulating current i_c. The two currents i_d'' and i_c during this period are represented by the simultaneous equations

$$[(2L_p+L_d)D+R]i_d''-L_pDi_c=e_A-e_C \qquad (2)$$

$$-L_pDi_d''+2L_pDi_c=e_B-e_A \qquad (3)$$

Figure 3 shows the 6-phase double-wye circuit with interphase transformer. The magnetizing current of the interphase

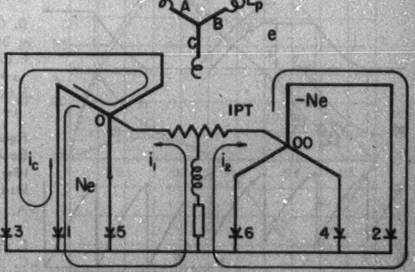


Figure 3. Diagram of 6-phase double-wye rectifier with interphase transformer IPT

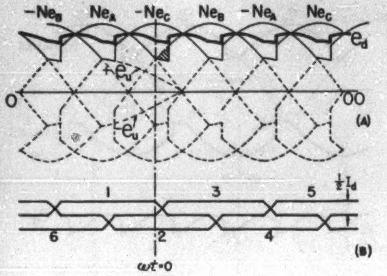


Figure 4. Voltage and current wave shapes of double-wye rectifier (characterized), first mode

A. Voltages
B. Currents

transformer is considered to be negligibly small, so that $i_1=i_2=i_d/2$.

During the period of single conduction

$$[(N^2L_p+L_d)D+R]i_1+[L_dD+R]i_2=Ne_A \qquad (4)$$

$$[L_pD+R]i_1+[(N^2L_p+L_d)D+R]i_2=-Ne_C \qquad (5)$$

Since $i_1+i_2=i_d$, equations 4 and 5 reduce to

$$[(N^2L_p+2L_d)D+2R]i_d=N(e_A-e_C) \qquad (6)$$

To make the two cases directly comparable, let $N=2$ and e, L_p, R, L_d be the same for both connections. Then equation 6 reduces exactly to equation 1, proving the 6-phase double-way and double-wye equivalent during the period of single conduction.

The period of "single" conduction ends and the overlap u begins with the pick-up of valve *3* when Ne_B becomes equal to $Ne_A-N^2L_pDi_1$. Remembering that $N=2$ and $i_1=i_d/2$, this condition again duplicates the performance of the 6-phase double-way rectifier. When $N=2$, $i_c'=i_c/2$. This is to be expected, since in the double-way circuit i_d is commutated in total, while in the double-wye circuit only $i_d/2$ is commutated at any one time. During the commutation period u

$$[(N^2L_p+L_d)D+R]i_1+[L_dD+R]i_2-N^2L_pDi_c'=Ne_A \qquad (7)$$

$$[L_dD+R]i_1+[(N^2L_p+L_d)D+R]i_2=-Ne_C \qquad (8)$$

$$N^2L_pDi_1+2N^2L_pDi_c'=N(e_B-e_A) \qquad (9)$$

Substituting $i_1+i_2=i_d$ and $N=2$ these equations reduce to

$$[(2L_p+L_d)D+R]i_d-2L_pDi_c'=e_A-e_C \qquad (10)$$

$$-L_pDi_d+4L_pDi_c'=e_B-e_C \qquad (11)$$

Substituting $i_c'=i_c/2$ these equations become identical to equations 2 and 3 for the double-way rectifier.

The value of i_d at the beginning of the period of single conduction must be the same as the value of i_d'' at the end of the preceding period of overlap. Under steady-state conditions, the value of i_d at the end of the period of single conduction must also be equal to the value of i_d'' at the beginning of the preceding period of overlap. These conditions uniquely determine the constants of integration of the preceding differential equations.

It can be seen in Figure 2(A) that during commutation between valves *6* and *2* of the double-way rectifier the terminal voltage of phase B is depressed along curve e_u and will remain negative with respect to the terminal voltage of phase A until the commutation *6-2* is completed. Valve *3* is therefore prevented from picking up at least until that time. Since this behavior is repeated at each of the six commutation periods per cycle, commutation is automatically limited to a maximum of 60 electrical degrees during the first and second modes of operation regardless of the load. When u reaches 60 degrees, a further increase of load current will cause a delay in pick-up resembling phase control.

The 6-phase double-wye circuit exhibits exactly the same property. Refer to Figures 4(A) and 4(B). During commutation between valves *6* and *2* the terminal voltage of d-c winding $-Ne_B$ with respect to neutral 00 is raised along curve e_u. This winding is closely coupled to winding $+Ne_B$ and their voltages must at

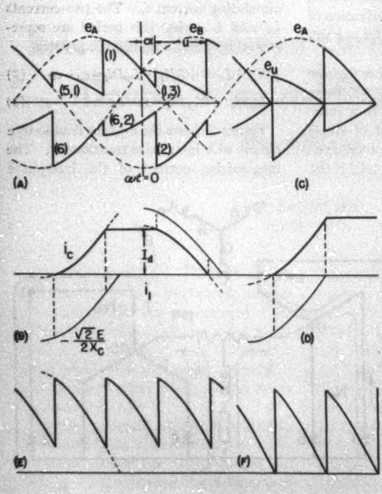
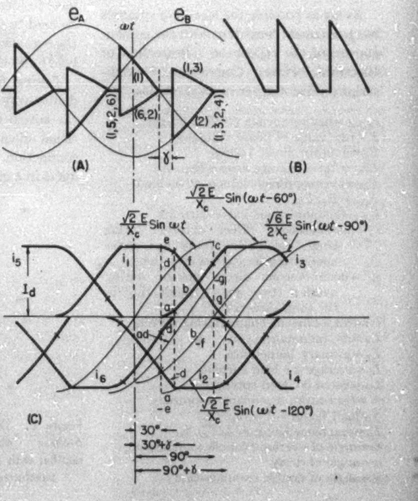


Figure 5 (left). Double-way rectifier, inductive load, second mode

A. Alternating voltages, $\alpha=15$ degrees
B. Current, $\alpha=15$ degrees
C. Alternating voltages, $\alpha=30$ degrees
D. Current, $\alpha=30$ degrees
E. Direct voltage, $\alpha=15$ degrees
F. Direct voltage, $\alpha=30$ degrees

Figure 6 (right). Double-way rectifier, inductive load, third mode

A. Alternating voltages
B. Direct voltage
C. Currents



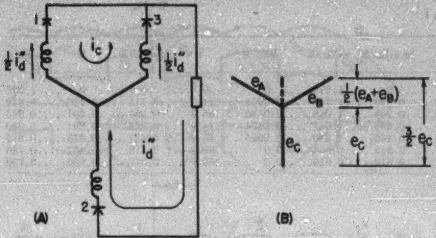


Figure 7 (above). (A). Functional diagram of double-way rectifier, nonincuctive load, during period of overlap. (B). Phasor diagram for this condition

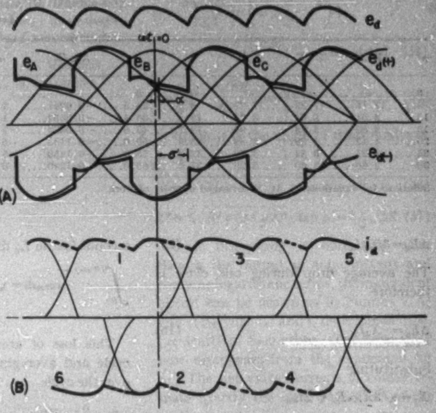


Figure 8 (right). Double-way rectifier, noninductive load, first mode

A. Voltages
B. Currents

all times be equal and opposite. Therefore, during this period, the terminal voltage of winding Ne_B with respect to neutral 0 is depressed along line $-e_u$ and will remain negative with respect to the terminal voltage of phase Ne_A until the commutation *6-2* is completed. Again, valve *3* is prevented from conducting until that time.

Similarly, it can be shown that both circuits behave the same way during periods of double commutation, that is when two commutations occur at the same time. The third mode of operation with d-c inductance involves double commutation and the identity of the two circuits will become obvious during the derivation of the equations.

The preceding discussion presents the proof of the identity of the behavior of the two connections and sets down the equations and limits required in the derivation of the complete regulation curves. We can proceed, then, with the derivations for the case of the double-way circuit and by means of per-unit quantities make the formulas directly applicable to both circuits.

Highly Inductive Load

FIRST MODE

At no load, there is of course no overlap and the direct voltage is the average difference between the positive and negative envelopes of the phase voltages.

$$E_{d0}=\frac{3}{1/f}\int_0^{1/2f}(e_B-e_A)dt=\frac{3}{\pi}\sqrt{6}E=2.34E \qquad (12)$$

In the case of the double-wye circuit, the instantaneous direct voltage is the average of the instantaneous direct voltages of the two 3-phase systems. From considerations of symmetry, the average direct voltages of the two systems are equal and therefore each equal to the average combined direct voltage. Then

$$E_{d0}=\frac{3\sqrt{3}\sqrt{2}NE}{2/f}=\int_0^{1/2f}\sin\omega t\,dt$$
$$=\frac{3}{\pi}\sqrt{\frac{3}{2}}NE=1.17NE \qquad (13)$$

which, again, becomes identical to equation 12 when $N=2$.

To determine the drop in direct voltage due to primary reactance, refer to Figure 2 and note that at each commutation we lose from the time integral of the direct voltage an area equal in volt-seconds to

$$\frac{\Delta E_d}{6f}=\int_a^b e_{\text{drop}}dt=n(\phi_b-\phi_a)10^{-8}$$
$$=L_p(i_b-i_a) \qquad (14)$$

a fundamental relationship in which i_a and i_b are the values of the current through a primary reactance at the beginning and end of a commutating period, and $n(\phi_b-\phi_a)$ is the equivalent change of flux linkages.

In the limiting case of infinite load inductance, the valve and phase currents are flat-topped at the value I_d and $i_b-i_a=I_d$. Therefore

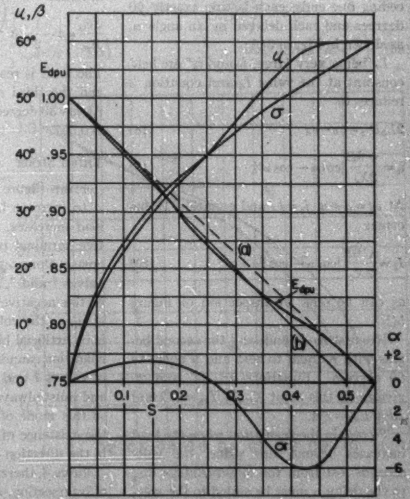


Figure 9. Noninductive load, first mode. E_{dpu}, u, α and β plotted against S. ($S=X_{cpu}I_d/I_{dn}$)

A. Plot of E_{dpu} if α remained zero

B. E_{dpu} for inductive load

Table I

ϕ''	R/X	Z''/X	Z'/X	ϕ'	σ	AX_c/E	α	$i_{d(\alpha)}X_c/E$	CX_c/E	BX_c/E	u	$i_{d(\sigma)}X_c/E$	S	E_{dpu}
0	∞	∞	∞	0	0	0	0	0	1.225	0	0	0	0	1.000
5	17.16	17.23	17.27	6.6	25.5	−0.0241	1.40	0.129	1.161	0.0058	24.1	0.115	0.106	0.950
10	8.51	8.64	8.74	13.2	36.4	−0.0584	1.10	0.258	1.096	0.0150	35.3	0.221	0.201	0.896
15	5.59	5.78	5.94	19.7	43.8	−0.0875	−1.07	0.366	1.042	0.0134	44.9	0.323	0.291	0.850
20	4.12	4.38	4.58	25.9	50.0	−0.1133	−5.23	0.440	1.000	0.0025	55.2	0.421	0.378	0.815
25	3.21	3.54	3.78	32.0	55.0	−0.1250	−5.00	0.519	0.966	0.0	60.0	0.519	0.467	0.786
30	2.60	3.00	3.28	37.6	60.0	−0.1290	0	0.612	0.919	0.0	60.0	0.612	0.550	0.750

Solutions for equations 33, 34, 35 arranged in order of work.

$$\Delta E_d = 6fL_pI_d \tag{15}$$

The average drop during one cycle is therefore

$$\Delta E_d = \frac{3}{\pi} X_c I_d \tag{16}$$

Substituting

$$X_c = \sqrt{3} X_{cpu} E / \sqrt{2} I_{dn}$$

and

$$E = \pi E_{d0} / 3\sqrt{6}$$

$$\frac{\Delta E_d}{E_{d0}} = \frac{1}{2} X_{cpu} \frac{I_d}{I_{dn}} = \frac{1}{2} S \tag{17}$$

Second Mode

Equation 17 holds from no load until the overlap u reaches 60 degrees. As the load is increased still further the terminal voltage of phase B remains depressed along e_u [Figure 2(A)], preventing the conduction of valve *3* until commutation *6-2* is completed and the terminal voltage returns to normal. This process is repeated at the end of each commutation period so that one begins where the preceding one ends, each lasting exactly 60 degrees and each delayed by an angle α, as shown in Figure 5(A).

L_d being very large, i_d and i_d'' are held constant at the value I_d and equation 3 reduces to

$$2L_pDi_c = e_B - e_A \tag{18}$$

$$i_c = \frac{\sqrt{6}E}{2X_c}[\cos\alpha - \cos\omega t] \tag{19}$$

At $\omega t = \alpha + u$, $i_c = I_d$ and equation 19 becomes

$$I_d = \frac{\sqrt{6}E}{2X_c}[\cos\alpha - \cos(\alpha + u)] \tag{20}$$

as can be seen by inspection of Figure 5(B).

The first mode ends and the second begins with α equal to zero and u equal to 60 degrees. Equations 16, 17, and 20 give us at this point $X_{cpu}I_d/I_{dn} = 1/2$ and $\Delta E_d/E_{d0} = 1/4$.

To obtain the regulation when the load increases beyond this value, the volt-seconds lost from the direct voltage due to the delay α must be added to the volt-seconds lost in L_p during commutation.

$$\int_0^{(\alpha+u)/\omega} e_{\text{drop}}dt = \sqrt{6}E \int_0^{\alpha/\omega} \sin\omega t dt + L_pI_d \tag{21}$$

This loss of area occurs six times a cycle and averaging the total lost area over the cycle

$$\Delta E_d = \frac{3}{\pi}[\sqrt{6}E(1 - \cos\alpha) + X_cI_d] \tag{22}$$

which reduces to

$$\frac{\Delta E_d}{E_{d0}} = (1 - \cos\alpha) + \frac{1}{2} S \tag{23}$$

Substituting for X_c and making $u = \pi/3$ in equation 20

$$S = \cos\alpha - \cos\left(\alpha + \frac{\pi}{3}\right)$$

or

$$\cos\alpha = \frac{1}{2} S - \frac{1}{2}\sqrt{3}\sqrt{1 - S^2} \tag{24}$$

Replacing $\cos\alpha$ in equation 23 the per-unit voltage drop during the second mode of operation with highly inductive load becomes

$$\frac{\Delta E_d}{E_{d0}} = 1 - \frac{1}{2}\sqrt{3}\sqrt{1 - S^2} \tag{25}$$

The limit is reached when $\alpha = 30$ degrees and the third mode of operation begins. At $\alpha = 30$ degrees, $S = \sqrt{3}/2$ and $\Delta E_d/E_{d0} = 0.567$.

Third Mode

From Figure 5(C) it is seen that were α to increase beyond 30 degrees as the load increases, the potential of the positive terminal of the rectifier would decrease along e_u, the average voltage of valves *1* and *3*, and fall below the voltage of the negative terminal which is rising along e_C, the voltage of valve *2*. As there is no artificial blocking of the valves, this condition cannot exist: The a-c terminals of valves *1* and *4* are connected together and must always have the same potential so this mode of operation would require the existence of a voltage across valve *4* in the direction of current flow.

Valve *4*, therefore, picks up at the point of intersection, and for a time γ all three phases are involved in commutation, the net output voltage being zero for this period.

At first glance it appears valves *5* and *6* should also pick up and carry current during this interval. Current wave shapes can be constructed for simultaneous conduction of all valves, which will satisfy the conditions of a perfectly smooth d-c output and zero valve voltages, but only at complete short circuit. However, simultaneous conduction of only valves *1*, *3*, *2*, and *4* satisfies these conditions while approaching complete short circuit and produces a lower value of steady-state short-circuit current. Even in an ideal circuit the higher short-circuit current could only be initiated by a transient condition.

Figure 6(A) shows the voltage conditions in the rectifier as short circuit is approached and Figure 6(B) the output voltage of the rectifier for the same conditions. Figure 6(C) demonstrates the construction of the valve currents. It should be noted here that each valve carries current one-half cycle plus γ.

During the overlap between valves *1* and *4*, the time rate of change of the sum of these two currents must equal the time rate of change of the short-circuit current of phase A. There are no losses in the path of the commutating current, so we can say that the sum of the sinusoidal components of these currents must equal the short-circuit current of phase A and the voltage during that period will then be zero.

The lettered sections of the current curves in Figure 6(C) refer to the sinusoidal components which are displaced by a suitable d-c component from similarly lettered sections on the light sinusoidal curves. Adding up the height of each section between the limits indicated below the figure, we have

$$I_d = \frac{\sqrt{2}E}{X_c}\sin\left(\omega t - \frac{2}{3}\pi\right)\Big|_{\pi/6}^{\pi/6+\gamma} + \frac{\sqrt{6}E}{2X_c}\sin\left(\omega t - \frac{\pi}{2}\right)\Big|_{\pi/6+\gamma}^{\pi/2} + \frac{\sqrt{2}E}{X_c}\sin\left(\omega t - \frac{\pi}{3}\right)\Big|_{\pi/2}^{\pi/2+\gamma} \tag{26}$$

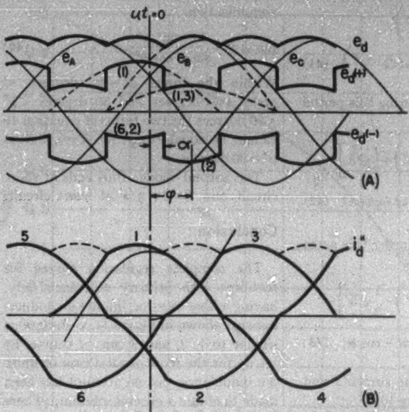


Figure 10. Double-way rectifier, noninductive load, second mode

A. Voltages
B. Currents

Evaluating the three terms and substituting for X_c

$$\sin\left(\gamma+\frac{\pi}{6}\right)=\sqrt{3}S-1 \qquad (27)$$

In evaluating the voltage drop during the third mode of operation, the reactance drop through the primary reactance is caused by the rise of current from 0 to I_d through, say valve *3*, and the final decay of current in valve *6* during the period γ.

$$\Delta E_d=6f\left[\frac{\sqrt{6}E}{\omega}\int_0^{\pi/6}\sin\omega t\,d\omega t+\frac{X_c}{w}\left\{I_d+\frac{\sqrt{2}E}{X_c}\sin\omega t\Big|_{\pi/6}^{\pi/6+\gamma}\right\}\right] \qquad (28)$$

With the help of equation 27, and making the necessary substitutions, this is reduced quite easily to

$$\frac{\Delta E_d}{E_{d0}}=1-\sqrt{3}+\frac{3}{2}S \qquad (29)$$

which is linear with respect to load.

At full short circuit $\Delta E_d/E_{d0}=1$, $S=2/\sqrt{3}$ and $\gamma=60$ degrees.

Noninductive Load

First Mode

Referring back to Figures 1(B) and 1(C), but eliminating the d-c inductance L_d, equations 1, 2, and 3 become

$$[2L_pD+R]i_d=\sqrt{6}E\sin\left(\omega t+\frac{2}{3}\pi\right) \qquad (30)$$

$$[2L_pD+R]i_d''-L_pDi_c=\sqrt{6}E\sin\left(\omega t+\frac{2}{3}\pi\right) \qquad (31)$$

$$-L_pDi_d''+2L_pDi_c=\sqrt{6}E\sin\omega t \qquad (32)$$

In this and what follows the single prime applies to quantities and functions during the period of single conduction and the double prime applies to the period of overlap u. Angles u and α are the same as in the treatment of the case for highly inductive load, α occurring here even in the first mode. σ is used for convenience to denote the start of the first single conduction period and is the correct sum of $u+\alpha$ in the step-by-step evaluation of the constants of integration in the solutions of the preceding equations.

The solutions for i_d, i_d'' and i_c take the form

$$i_d=\frac{\sqrt{6}E}{Z'}\sin\left(\omega t+\frac{2}{3}\pi-\phi'\right)+A\epsilon^{-\frac{R}{2X_c}\left(\omega t+\frac{\pi}{3}-\sigma\right)}\left(\sigma-\frac{\pi}{3}<\omega t<\alpha\right) \qquad (33)$$

$$i_d''=\frac{3\sqrt{2}E}{2Z''}\sin\left(\omega t+\frac{\pi}{2}-\phi''\right)+B\epsilon^{-\frac{2R}{3X_c}(\omega t-\alpha)} \quad (\alpha<\omega t<\sigma) \qquad (34)$$

$$i_c=\frac{1}{2}i_d''-\frac{\sqrt{6}E}{2X_c}\cos\omega t+C \quad (\alpha<\omega t<\sigma) \qquad (35)$$

in which α is the angle of delay and $\sigma=\alpha+u$

$$Z'=\sqrt{R^2+(2X_c)^2},\qquad \tan\phi'=2X_c/R \qquad (36)$$

$$Z''=\sqrt{R^2+(3X_c/2)^2},\ \tan\phi''=3X_c/2R \qquad (37)$$

It is interesting to note that even though the output is not constant and contains logarithmic terms, the equivalent circuit can be drawn as in Figures 7(A) and 7(B), with i_d'' and i_c determined independently of each other. The term $i_d''/2$ now disappears from the expression for i_c. The net valve currents of course remain exactly the same regardless of the method used.

Returning to equations 34 and 35, it is necessary to determine α, the point at which overlap begins. This can occur only where the terminal voltage of the rectifier intersects the terminal voltage of the next phase to conduct current as shown in Figures 8(A) and 8(B). In this the first mode of operation there is no current through L_p of that phase so that its terminal and induced voltages are equal. Except that u is not necessarily 60 degrees, the action is very similar to the behavior of the rectifier with d-c inductance during its second mode of operation. The necessary condition is expressed by

$$\sqrt{2}E\sin\left(\omega t+\frac{5}{6}\pi\right)-L_pDi_d=\sqrt{2}E\sin\left(\omega t+\frac{\pi}{6}\right) \qquad (38)$$

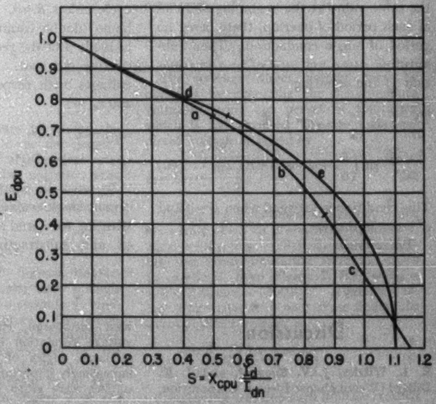


Figure 11. Regulation curves, no load to short circuit

A. Inductive load, first mode
B. Inductive load, second mode
C. Inductive load, third mode
D. Noninductive load, first mode
E. Noninductive load, second mode

which equation is satisfied at $\omega t=\alpha$. Making the necessary substitutions we obtain

$$\sin\alpha=\frac{X_c}{Z'}\sin\left(\frac{\pi}{6}+\alpha-\phi'\right)+\frac{AR}{2\sqrt{6}E}\epsilon^{-\frac{R}{2X_c}\left(\frac{\pi}{3}-u\right)} \qquad (39)$$

The constants of integrations A, B, and C and the angles α and u are best determined by successive approximations based on the necessary conditions that i_d at the beginning of single conduction is equal to i_d'' at the end of overlap, and i_d at the end of single conduction is equal to i_d'' at the beginning of overlap.

The average direct current I_d is obtained by integration of i_d and i_d'' (after substitution of constants in equations 33 and 34) over the limits already indicated and averaging over one-sixth of a cycle. The average direct voltage is then RI_d.

The results are given in Table I arranged in 5-degree intervals of ϕ''. E_{dpu}, u, and α are plotted against S in Figure 9. E_{dpu} for an inductive load is also shown for comparison. The dotted line in this figure shows what the output voltage would be for a noninductive load if A and B of equations 33 and 34 were assumed to be zero throughout this mode of operation.

One of the surprising results is the inversion of α during the latter half of the second mode; that is, the start of commutation before the intersection of phase voltages. Also to be noted is the fact that u becomes practically 60 degrees considerably before the end of this mode of operation.

SECOND MODE

Since u must be equal to 60 degrees during the second mode, i_d'' must have the same value at the beginning and end of each period of overlap, there being no period of single conduction. Then substituting $\omega t=\alpha$ and $\omega t=\pi/3+\alpha$ in equation 34 and equating

$$\frac{3\sqrt{2}E}{2Z''}\sin\left(\frac{\pi}{2}+\alpha-\phi''\right)+B=\frac{3\sqrt{2}E}{2Z''}\sin\left(\frac{5}{6}\pi+\alpha-\phi''\right)+B\epsilon^{-\frac{2R}{3X_c}\frac{\pi}{3}} \qquad (40)$$

This condition is satisfied when $B=0$ and $\phi''-\alpha=30$ degrees.

Therefore

$$i_d''=\frac{3\sqrt{2}E}{2Z''}\sin\left(\omega t+\frac{\pi}{2}-\phi''\right)\quad\left(\phi''-\frac{\pi}{6}<\omega t<\phi''+\frac{\pi}{6}\right) \qquad (41)$$

The current in valve *1* during this period is

$$i_1=\frac{3\sqrt{2}E}{4Z''}\left[\sin\left(\omega t+\frac{\pi}{2}-\phi''\right)+\frac{1}{2}\sqrt{3}\right]+\frac{\sqrt{6}E}{2X_c}[\cos\omega t-\cos\alpha] \qquad (42)$$

The current in valve *3* during this same period is

$$i_3=\frac{3\sqrt{2}E}{4Z''}\left[\sin\left(\omega t+\frac{\pi}{2}-\phi''\right)-\frac{1}{2}\sqrt{3}\right]-\frac{\sqrt{6}E}{2X_c}[\cos\omega t-\cos\alpha] \qquad (43)$$

The construction of the currents and voltages is shown in Figure 10.

The average direct current is obtained by integrating and averaging i_d'' over the interval $\phi''-\pi/6$ to $\phi''+\pi/6$.

$$I_d=\frac{6f}{\omega}\int_{\phi''-\pi/6}^{\phi''+\pi/6} i_d''\,d\omega t=\frac{3}{\pi}\frac{3}{2}\frac{\sqrt{2}E}{Z''} \qquad (44)$$

Multiplying by $\sqrt{R^2+(3X_c/2)^2}=Z''$

$$\sqrt{(RI_d)^2+\left(\frac{3}{2}X_cI_d\right)^2}=\frac{3}{\pi}\frac{3}{2}\sqrt{2}E \qquad (45)$$

$RI_d=E_d$. Substituting for RI_d and X_c in equation 45 and dividing by E_{d0} we obtain by simple manipulation

$$E_{dpu}=\frac{E_d}{E_{d0}}=\frac{1}{2}\sqrt{3}\sqrt{1-\left(\frac{\pi}{2\sqrt{3}}S\right)^2} \qquad (46)$$

From equations 41, 42, and 43 and Figure 10 it is obvious that the instantaneous value of the direct current can never fall to zero in the second mode of operation and the terminal voltage of the rectifier cannot be zero during any part of the cycle unless $R=0$. Therefore, there can be no "double commutation" as occurred in mode *3* of the rectifier with highly inductive d-c load. Calling the terminal voltages with respect to neutral $e_d(+)$ and $e_d(-)$

$$e_{d(+)}=e_A-L_pDi_1 \qquad (47)$$

$$e_{d(-)}=e_C+L_pDi_d'' \qquad (48)$$

It must be shown that $e_{d(+)}-e_{d(-)}$ cannot be zero except when $R=0$. Substituting for i_1 and i_d'' in equations 47 and 48 and substracting we obtain after simplification

$$\tan\omega t\neq-\frac{1-\sin^2\phi''}{\sin\phi''\cos\phi''} \qquad (49)$$

Within the limits $\phi''-30$ degrees $<\omega t<\phi''+30$ degrees and 30 degrees $<\phi''<90$ degrees the two terms of equation 49 cannot be equal except for $R=0$, that is $\phi''=90$ degrees.

The second mode, thus ends at short circuit and $S=2\sqrt{3}/\pi$ at short circuit.

Conclusion

The complete regulation curves for rectifiers with primary reactance only, having either zero or infinite d-c inductance are shown in Figure 11. Equations similar to 1, 2, and 3 can, of course, be set up for the more general case of finite d-c inductance, but no attempt has been made to obtain a general solution. There is no reason to believe the curves will not fall between the two limiting cases presented here.

The case for mixed primary and secondary reactance presents much greater difficulty and from the work done so far it appears that specific solutions for particular problems will be in order for some time unless investigation is being carried on in other quarters.

As previously mentioned, the author's earlier paper[4] contained two errors which can now be corrected. The first mode of operation for the case with noninductive d-c load was given as a straight line; the resulting error in computation is not great. The second error occurred in the omission of mode *3* in the case of a rectifier with highly inductive d-c load. Short-circuit computations based on that paper will probable remain valid since losses in the rectifier unit depress the regulation curve so that the third mode of operation is almost entirely eliminated.

References

1. DIE STROM UND SPANNUNGSVERHAELTNISSE DER GROSSGLEICHRICHTER, **Walter Daellenbach, E. Gerecke.** *Archiv fu:r Elektrotechnik* (Berlin, Germany), volume 14, number 2, 1924.

2. MERCURY ARC POWER RECTIFIERS (book), **Othmar K. Marti, Harold Winograd.** McGraw Hill Book Company, Inc., New York, N. Y., 1930.

3. PROTECTION OF ELECTRONIC POWER CONVERTERS, **AIEE Committee Report.** *AIEE Transactions,* volume 69, part II, 1950, pages 813–29.

4. EQUIVALENT MACHINE CONSTANTS FOR RECTIFIERS, **I. K. Dortort.** *Electrical Engineering,* volume 70, September 1951, page 802.

◆

Discussion

R. L. Witzke, J. V. Kresser, and **J. K. Dillard** (Westinghouse Electric Corporation, East Pittsburgh, Pa.): The question of voltage regulation of rectifiers operating at large angles of overlap is becoming increasingly important; two papers on this subject were given at the meeting. Mr. Dortort's paper is more general since it covers the case of a noninductive as well as an inductive load. Taking slightly different analytical approaches, both papers reach the same conclusions with respect to

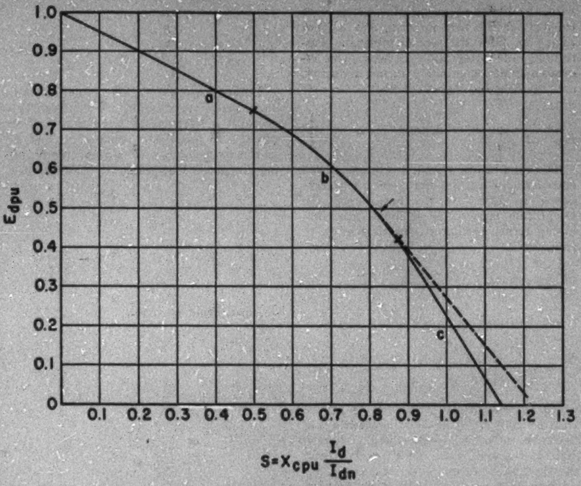


Figure 1. Voltage regulation curve showing effect of leakage between transformer secondary windings

—— $X_{ss} = 0$
- - - - $X_{ss} = 0.25\ X_c$

cases treated in both papers, and our paper[1] contains laboratory data to substantiate the calculated curves. Neither paper treats extensively the case for mixed primary and secondary reactance which occurs where the leakage reactance between the secondary windings of the double-wye transformer is not negligible. Knowledge of the effect of leakage reactance is essential, however, if the curves based on negligible leakage are to be used with confidence in practical cases.

The regulation curve for inductive load from Figure 11 of the Dortort paper, derived for zero reactance between secondary windings, is reproduced by the solid line of Figure 1 of the discussion. The first mode of operation is along curve *a* out to $S = 0.500$; the second mode is along curve *b* out to $S = 0.866$; the third mode is along curve *c* out to short circuit at $S = 1.15$. Taking leakage reactance into account, equations for the regulation curve are derived and given later. The first and second modes of operation are the same as for the case of negligible leakage reactance X_{ss}, except that the second mode terminates not at 30 degrees delay but at some smaller angle determined by equation 14 of the discussion. In general, the reactance between secondary windings on the same core will be in the range of 0.5 to 2.0 per cent on the a-c winding kilovolt-ampere base.[2] Taking the upper limit, the leakage reactance might be of the order of 25 per cent of the commutating reactance. In this case, $\alpha_m = 25.9$ degrees and the second mode terminates at $S = 0.827$, the point designated by the arrow in Figure 1 of the discussion. The third mode of operation exists along the dashed curve of the figure, rather than along curve *c*, ending just short of short circuit at $S = 1.21$. The lower limit of leakage reactance would give a curve between the solid and dashed lines of Figure 1. There follows a fourth mode of operation, which does not exist where $X_{ss} = 0$, during which simultaneous commutations occur within the same wye. Equations for the fourth mode were not derived since, for practical limits of leakage reactance, the third mode of operation carries almost to short circuit.

One other point is of interest. If X_{ss} is negligible, $\alpha_m = 30$ degrees and equations 22 and 24 of the derivations given here reduce to

$$\frac{I_c X_c}{E_s} = \frac{\sqrt{2}}{2} + \frac{\sqrt{6}}{4} \sin u - \frac{\sqrt{2}}{4} \cos u$$

$$\frac{E_d}{E_{d0}} = \frac{\sqrt{3}}{2}\, 1 - \sin(u - 30°)$$

These are the equations given for the third mode of operation in our paper.[1] Therefore, if the reactance between secondary windings can be neglected, the operation of a 6-phase double-wye rectifier is identical to that of a 6-phase double-way rectifier.

DERIVATION OF EQUATIONS

Transformers used in 6-phase double-wye rectifier circuits usually have a single primary winding and two secondary windings displaced by 180 degrees. The circuit is easier to analyze if the transformer leakage reactance is represented by the conventional equivalent circuit of a 3-winding transformer external to an ideal transformer as shown in Figure 2 of the discussion. Ideal transformers permit the use of two separate primary windings. The reactances X_{ps} and X_{ss} are the actual leakage reactances between primary and each secondary winding and between secondary windings respectively. The turns ratio is taken as $\sqrt{3}$ in order to make the ratio of transformation between primary and secondary line-to-line voltages unity. The derivations are based on a highly inductive load.

Line-to-neutral voltages behind source reactance are E_a, E_b, E_c, defined by the phases of Figure 2(B). Commutation from phase 1 to phase 3 occurs when $E_1 = E_3$. If $E_a = \sqrt{2}\ E_s \cos(\omega t + \pi/2)$, where E_s is the rms value of transformer secondary no-load line-to-neutral voltage, commutation from phase 1 to phase 3 normally starts when $t = 0$. This fact may be deduced from the phase diagram of Figure 2 of the discussion, which takes cognizance of the phase shift through the delta-wye transformer bank.

FIRST MODE OF OPERATION

During a phase 1 to phase 3 commutation, if $L_c = L_s + L_{ps}$ and $p = d/dt$

$$E_{ab} = -\frac{3}{2} L_c p i_1 + E_{a'b'}$$

$$E_{bc} = E_{b'c'} \tag{1}$$

$$E_{a'b'} = E_{c'a'} = -\frac{E_{bc}}{2}$$

The last equation arises from the consideration that $E_3 = E_1$ and the sum of the voltages around the delta is zero.

The solution of equation 1 gives

$$p i_1 = -\left[\frac{2E_{ab} + E_{bc}}{3L_c}\right] = -\frac{E_a}{L_c} \tag{2}$$

The commutating angle is obtained by integrating equation 2 over the commutating interval and evaluating the integral at the boundaries $\omega t = 0$ at which the current $i = 0$ and $\omega t = u$ at which $i = I_d/\sqrt{3}$. The latter value is obtained from the fact that $i_{cc} = i\sqrt{3}/2$ and the final value of $i_{cc} = I_d/2$. This is shown in Figure 2 of the discussion. After integrating and rearranging

$$\cos u = 1 - \frac{I_d X_c}{\sqrt{6} E_s} \tag{3}$$

The average direct voltage may be obtained by averaging the instantaneous voltages during the period of symmetry which is 120 degrees. The instantaneous direct voltage is given by

$$e_d = E_3 = -\frac{E_{a'b'}}{\sqrt{3}} \tag{4}$$

During the commutation interval, which lasts from $\omega t = 0$ to $\omega t = u$, from equation 1, $E_{a'b'} = -E_{bc}/2 = -\sqrt{3}/2\ E_s \cos \omega t$. During the noncommutation interval, $\omega t = u$ to $\omega t = 2\pi/3$, from Figure 2, $E_{a'b'} = E_{ab} = \sqrt{3}\ \sqrt{2}\ E_s \cos(\omega t + 2\pi/3)$. Therefore, integrating equation 4 over these limits, the average direct voltage is

$$E_d = \frac{3\sqrt{6}}{2\pi} E_s \left(\frac{1 + \cos u}{2}\right) \tag{5}$$

Neglecting arc drop each 3-phase rectifier group has this average output voltage which is also the average direct voltage across the load. The system is equivalent to two 3-phase rectifiers in parallel with 60 degrees shift between the voltages, the difference between the two sets of instantaneous voltages being absorbed by the interphase transformer.

The no-load direct voltage E_{d0} is obtained from equation 5 with $u=0$. Substituting 3 into 5 and dividing by E_{d0}

$$\frac{E_d}{E_{d0}}=1-\frac{I_dX_c}{2\sqrt{6}E_s} \tag{6}$$

For the double-wye circuit $X_c=\dfrac{\sqrt{6}E_s}{I_{dn}}X_{cpu}$.

Therefore

$$\frac{E_d}{E_{d0}}=1-\frac{1}{2}\frac{I_d}{I_{dn}}X_{cpu}=1-\frac{1}{2}S \tag{7}$$

which corresponds to equation 17 of the Dortort paper. This shows that operation during the first mode is unaffected by leakage reactance between the secondary windings.

SECOND MODE OF OPERATION

As Mr. Dortort points out, and as also derived in our paper,[1] when the overlap angle reaches 60 degrees, commutations are delayed by an angle α. For this condition, the average direct voltage is obtained from equation 5 of the discussion, with the limits of integration being α to $\alpha+\pi/3$ for the first integral, and $\alpha+\pi/3$ to $\alpha+2\pi/3$ for the second integral. Performing the integration between the indicated limits, combining terms, reducing, and dividing by $E_{d0}=3\sqrt{6}\,E_s/2\pi$

$$\frac{E_d}{E_{d0}}=\frac{\sqrt{3}}{2}\cos(\alpha+30\text{ degrees}) \tag{8}$$

The commutating angle is obtained from equation 2 with the limits of integration $\omega t=\alpha$ to $\omega t=\alpha+\pi/3$. Performing the integration between these limits and combining terms, similar to the derivation of equation 3

$$\frac{I_dX_c}{E_s}=\sqrt{6}\sin\,(\alpha+30\text{ degrees}) \tag{9}$$

Substituting equation 9 into 8

$$\frac{E_d}{E_{d0}}=\sqrt{\frac{3}{4}-\frac{1}{8}\left[\frac{I_dX_c}{E_s}\right]^2}=\frac{\sqrt{3}}{2}\sqrt{1-S^2} \tag{10}$$

which corresponds to equation 25 of the Dortort paper. If leakage reactance between secondary windings is negligible, the limit is reached when $\alpha=30$ degrees and the third mode of operation begins. Considering leakage, the limit is reached at $\alpha<30$ degrees as shown now.

Following the same reasoning used in writing equation 1, the voltages in the other delta during a phase 1 to phase 3 commutation are found to be

$$E_{bc}=E_{b''c''}$$

$$E_{ca}=\frac{3}{2}\left(L_c-\frac{L_{ss}}{2}\right)pi_1+E_{c''a''} \tag{11}$$

From Figure 2 of the discussion it can be seen that commutation from phase 1 to phase 3 is followed by commutation from phase 2 to phase 4. Defining E_2' and E_4' as the values of E_2 and E_4 respectively during the phase 1 to phase 3 commutation

$$E_2'=\frac{E_{b''c''}}{\sqrt{3}}=\frac{E_{bc}}{\sqrt{3}}$$

$$E_4'=\frac{E_{c''a''}}{\sqrt{3}}=\frac{1}{\sqrt{3}}\left[E_{ca}-\frac{3}{2}\left(L_c-\frac{L_{ss}}{2}\right)pi_1\right] \tag{12}$$

Substituting equation 2 into 12

$$E_2'=\frac{E_{bc}}{\sqrt{3}}$$

$$E_4'=\frac{1}{\sqrt{3}}\left[E_{ca}+\frac{3}{2}\left(\frac{2X_c-X_{ss}}{2X_c}\right)E_a\right]$$

$$=\frac{1}{\sqrt{3}}\left[E_{ca}+\frac{6X_c-3X_{ss}}{4X_c}E_a\right] \tag{13}$$

No-load values of these voltages are $E_2=E_{bc}/\sqrt{3}$ and $E_4=E_{ca}/\sqrt{3}$. Therefore, E_2 is unchanged but E_4 is reduced in magnitude and shifted back in phase angle while there is a phase 1 to phase 3 commutation in the other wye.

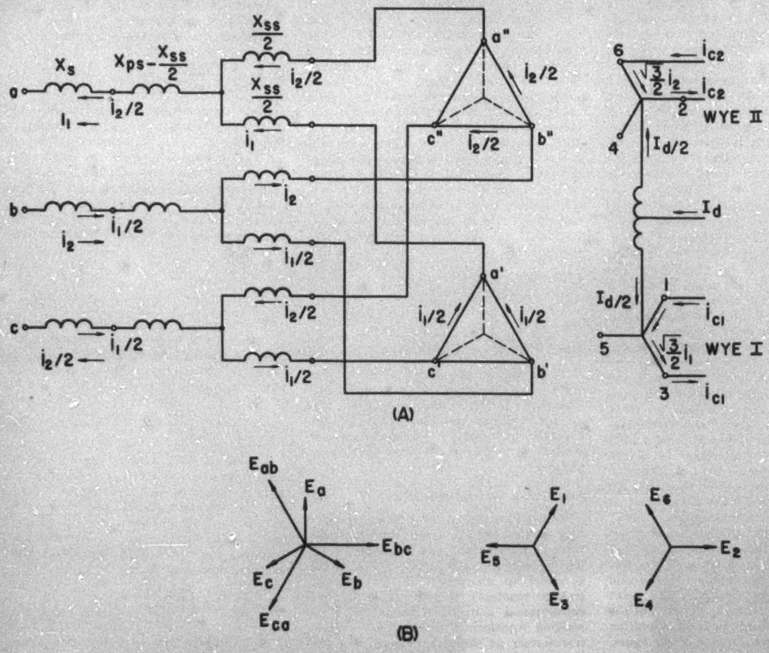


Figure 2 (A). Circuit diagram for double-wye circuit. (B) Voltage phase diagrams at no load

The maximum value of α for which equation 10 is valid is determined from 13 by equating E_2' and E_4' for the instant $\omega t=\pi/3+\alpha_m$ since the normal start of commutation is $\omega t=\pi/3$. Rewriting the voltages of 13 in accordance with the phase diagram of Figure 2, substituting $\omega t=\pi/3+\alpha_m$, equating the two equations of 13, expanding, and collecting terms

$$\tan\alpha_m=\left[\frac{2X_c-X_{ss}}{6X_c+X_{ss}}\right]\sqrt{3} \tag{14}$$

where X_c is the commutating reactance and X_{ss} is the leakage reactance between secondary windings.

Third Mode of Operation

When the forced delay angle reaches α_m as determined by equation 14 of the discussion, commutations will start in the second wye since the voltages E_2' and E_4' become equal. Hence, there will be a time when both wyes will commutate simultaneously. Because of the symmetry of operation, the simultaneous commutation interval δ is the same at the end as at the start of the commutating period. Since a new commutation starts every 60 degrees, the total commutating period is 60 degrees+δ, as shown in Figure 3 of the discussion. However, in each wye commutations must start 120 degrees apart, making the angle of symmetry 120 degrees. This mode of operation continues until $\delta=60$ degrees.

The 120-degree angle of symmetry is composed of four distinct intervals as shown in Figure 3. During the first interval, $\omega t=\alpha_m$ to $\omega t=\alpha_m+\delta$, simultaneous commutations exist from phase 6 to 2 and from phase 1 to 3. In the second interval, $\omega t=\alpha_m+\delta$ to $\omega t=\alpha_m+60$ degrees, the only commutation is from phase 1 to 3. During the third interval, $\omega t=\alpha_m+60$ degrees to $\omega t=\alpha_m+\delta+60$ degrees, there are simultaneous commutations from phase 1 to 3 and from phase 2 to 4. In the fourth interval, $\omega t=\alpha_m+\delta+60$ degrees to $\omega t=\alpha_m+120$ degrees, the only commutation is from phase 2 to 4.

Referring to Figure 2, the current i_2 is shown for the phase 6 to 2 commutation. For the phase 2 to 4 commutation, i_2 flows into terminal c and $i_2/2$ flows out terminals a and b. Bearing this fact in mind, the following equations apply for the 120-degree period beginning at the start of the phase 1 to 3 commutation illustrated in Figure 3

First Interval, Where $L_c=L_s+L_{ps}$ and $L=L_c-L_{ss}/2$

$$E_{ab}=-\frac{3}{2}L_cpi_1-\frac{3}{2}Lpi_2+E_{a'b'}=-\frac{3}{2}Lpi_1-\frac{3}{2}L_cpi_2+E_{a''b''}$$

$$E_{bc}=\frac{3}{2}pi_2+E_{b'c'}=\frac{3}{2}L_cpi_2+E_{b''c''} \tag{15}$$

$$E_{ca}=\frac{3}{2}L_cpi_1+E_{c'a'}=\frac{3}{2}Lpi_1+E_{c''a''}$$

$E_{a'b'}=E_{c'a'}=-\frac{1}{2}E_{b'c'}$ and

$$E_{a''b''}=E_{b''c''}=-\frac{1}{2}E_{c''a''}$$

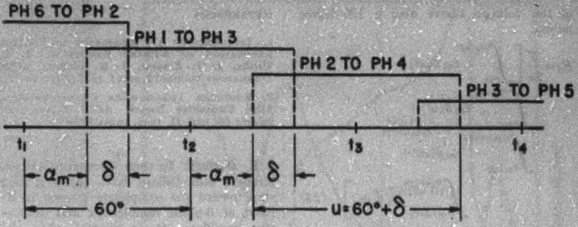


Figure 3. Graph showing interval of commutations in the third mode of operation

Solving equation 15

$$pi_1=-2\left[\frac{2L_cE_a+LE_b}{4L_c^2-L^2}\right]$$

$$pi_2=2\left[\frac{LE_a+2L_cE_b}{4L_c^2-L^2}\right] \tag{16}$$

Second Interval

The equations are given by 1 and the solution by 2.

$$pi_1=-\frac{2E_{ab}+E_{bc}}{3L_c}=-\frac{E_a}{L_c}$$

$$pi_2=0 \tag{17}$$

Third Interval

$$E_{ab}=-\frac{3}{2}L_cpi_1+E_{a'b'}=-\frac{3}{2}Lpi_1+E_{a''b''}$$

$$E_{bc}=-\frac{3}{2}Lpi_2+E_{b'c'}=-\frac{3}{2}L_cpi_2+E_{b''c''} \tag{18}$$

$$E_{ca}=\frac{3}{2}L_cpi_1+\frac{3}{2}Lpi_2+E_{c'a'}=\frac{3}{2}Lpi_1+\frac{3}{2}L_cpi_2+E_{c''a''}$$

$E_{a'b'}=E_{c'a'}=-\frac{1}{2}E_{b'c'}$ and

$$E_{b''c''}=E_{c''a''}=-\frac{1}{2}E_{a''b''}$$

Solving equation 18

$$pi_1=-2\left[\frac{2L_cE_a+LE_c}{4L_c^2-L^2}\right]$$

$$pi_2=2\left[\frac{LE_a+2L_cE_c}{4L_c^2-L^2}\right] \tag{19}$$

Fourth Interval

In this interval the only commutation is from phase 2 to 4. By following the same procedure used in obtaining equation 2

$$pi_1=0$$

$$pi_2=\frac{E_c}{L_c} \tag{20}$$

The instantaneous commutation current, $i_{c1}=\sqrt{3}\ i_1/2$, increases from zero at the start of the first interval to its final value, $I_c=I_d/2$, at the end of the third interval. Using equations 16, 17, and 19 and integrating between the limits of the intervals

$$I_c=-\frac{\sqrt{3}}{2}\left[\int_{\alpha m}^{\alpha_m+\delta}\frac{4X_cE_a+2XE_b}{4X_c^2-X^2}d(\omega t)+\int_{\alpha_m+\delta}^{\alpha_m+60}\frac{E_a}{X_c}d(\omega t)+\int_{\alpha_m+60}^{\alpha_m+\delta+60}\frac{4X_cE_a+2XE_c}{4X_c^2-X^2}d(\omega t)\right] \tag{21}$$

where $X_c=X_s+X_{ps}$ and $X=X_c-X_{ss}/2$.

As shown on Figure 2, the reference voltage $E_a=\sqrt{2}E_s\cos(\omega t+90\text{ degrees})$ and from Figure 3, $\delta=u-60$ degrees. Evaluating the integrals, collecting terms, combining, and multiplying by X_c/E_s

$$\frac{I_cX_c}{E_s}=\left[\sqrt{6}\cos\left(\frac{u}{2}+30\text{ degrees}\right)+\frac{6\sqrt{2}X_c}{2X_c+X}\sin\left(\frac{u}{2}-40\text{ degrees}\right)\right]\times\sin\left(\frac{u}{2}+\alpha_m\right) \tag{22}$$

The instantaneous direct voltage is obtained from equation 4 with $E_{a'b'}$ determined from the previously given circuit equations for each particular interval. This will give the following for the successive intervals:

From equations 15 and 16

$$e_{d1}=-\frac{E_{a'b'}}{\sqrt{3}}=\frac{1}{\sqrt{3}}\times\left[\frac{2(L_c^2-L^2)E_a+(4L_c^2-3L_cL-L^2)E_b}{4L_c^2-L^2}\right]$$

From 17 which came from 1

$$e_{d2}=-\frac{E_{a'b'}}{\sqrt{3}}=\frac{E_{bc}}{2\sqrt{3}}$$

From 18 and 19

$$e_{d3}=-\frac{E_{a'b'}}{\sqrt{3}}=\frac{1}{\sqrt{3}}\times\left[\frac{2(L_c^2-L^2)E_a+(4L_c^2-3L_cL-L^2)E_c}{4L_c^2-L^2}\right]$$

From equation 20 and Figure 2

$$e_{d4}=-\frac{E_{a'b'}}{\sqrt{3}}=-\frac{E_{ab}}{\sqrt{3}}$$

The average direct voltage is the average of the voltages above over a 120-degree period.

$$E_d=\frac{3}{2\pi}\left[\int_{\alpha_m}^{\alpha_m+\delta} e_{d1}\,d(\omega t)+\int_{\alpha_m+}^{\alpha_m+60} e_{d2}\,d(\omega t)+\int_{\alpha_m+60}^{\alpha_m+\delta+60} e_{d3}\,d(\omega t)+\int_{\alpha_m+\delta+60}^{\alpha_m+120} e_{d4}\,d(\omega t)\right] \quad (23)$$

Evaluating the integrals with $\delta=u-60$ degrees, collecting terms, reducing, and dividing by $E_{d0}=3\sqrt{6}/2\pi E_s$

$$\frac{E_d}{E_{d0}}=\sqrt{3}\cos\left(\frac{u}{2}+30\right)\cos\left(\frac{u}{2}+\alpha_m\right)+\frac{2}{\sqrt{3}}\left[\frac{X_c-X}{2X_c+X}\right]\sin\left(\frac{u}{2}+\alpha_m\right)\times\sin\left(\frac{u}{2}-30\right) \quad (24)$$

Operation during the third mode is given by equations 22 and 24 where α_m is determined by equation 14.

REFERENCES

1. INFLUENCE OF A-C REACTANCE ON VOLTAGE REGULATION OF 6-PHASE RECTIFIERS, R. L. Witzke, J. V. Kresser, J. K. Dillard. *AIEE Transactions*, volume 72, part I, 1953 (*Paper 53-54*).

2. RECTIFIER TRANSFORMER CHARACTERISTICS, AIEE Committee Report. *AIEE Transactions*, volume 68, part II, 1949, pages 1305–11.

I. K. Dortort: In their discussion Witzke, Kresser, and Dillard have taken a long step toward the completely general treatment of 6-phase double-way and double-wye rectifiers. It is evident that the subject of mixed primary and secondary reactances received a great deal of thought during the preparation of their own paper.[1]

Equation 14 of their discussion

$$\tan\alpha_m=\left[\frac{2X_c-X_{ss}}{6X_c+X_{ss}}\right]\sqrt{3}$$

defines a point of particular interest. When $X_{ss}=2X_c$ all primary reactance has been eliminated and the two 3-phase systems are completely isolated from each other; α_m becomes zero. The second mode of operation disappears and the first mode continues until

$$E_d/E_{d0}=1/2(\sqrt{3}-1) \text{ and } S=3-\sqrt{3}$$

If the current increases still further another mode of operation occurs, described in the early literature on rectifiers. The overlap increases until each anode conducts a full 360 degrees. The regulation curve is another straight line reaching short circuit at $S=2\sqrt{3}$.

The third mode is presented in the discussion as a straight line although this is not obvious from equation 24. It will be interesting (academically) to see if this mode stretches out as primary reactance decreases until it is transformed into the second mode of the classical 3-phase case. Or perhaps the second and third modes of mixed case both shrink and are replaced by the fourth mode (not described) which then transforms into the second mode of the classical case.

Of more practical interest, work remains to include the effect of losses, finite inductance in the load circuit, back electromotive force, and particularly the response of the rectifier under transient conditions.

REFERENCE

1. See reference 1 of Witzke, Kresser, and Dillard discussion.

Nondestructive Testing of Insulation

EMANUEL L. BRANCATO
MEMBER AIEE

THE USER of electrical apparatus is constantly in need of information concerning the condition of the insulation. This information, at the present time, is supplied by such tests as insulation resistance, power factor, capacitance, or by more severe tests such as "high-pot." Unfortunately, none of these tests give a true indication of the "health" of the dielectric but rather indicate its present condition—a condition that may exist only temporarily and can be corrected by suitable action. It is, of course, important to have this information; however, it is of equal importance in many applications to have an indication of the useful life than can be expected from the dielectric at any time during its use.

The user, as well as the designer, of electrical apparatus can predict the minimum useful life of the insulation in such equipment when operated under an assumed set of conditions by referring to appropriate thermal aging curves for the insulation. However, after the equipment has been in operation for a period of time, the post operation will influence the future life and these thermal aging curves are no longer useful. It becomes necessary then to be able to measure some property of the dielectric that is related to this thermal aging process and which takes into consideration the deterioration.

Several methods of nondestructive testing have been outlined or proposed;[1,2] however, for reasons of impracticability, or lack of sufficient evidence of reliability, these cannot be adopted. During early studies on various insulation measurements the electrical properties revealed a correlation with thermal aging. Further studies were then made on Formex-insulated magnet wire to verify these observations. These results are reported herein.

Experimental Procedure

Based on modern concepts of dielectrics[3-8] instrumentation was set up to measure capacitance, dielectric losses, and insulation resistance at suitable intervals during an aging cycle. A General Radio Company Schering Bridge, type *716-B* was employed in the measurement of capacitance and dielectric loss. This bridge is direct reading in capacitance at all audio frequencies, and in dissipation factor at frequencies of 1 kilocycle.

The bridge error signal is amplified by a General Radio Company type *814* amplifier to increase the balancing sensitivity. The balance is achieved on the fundamental by means of an appropriate filter. Detection of the unbalance is obtained with a Dumont *304H* cathode-ray oscilloscope. The null is sharply defined by a Lissajous pattern formed by the error and generated signals. The over-all sensitivity permits detection to within 0.02 per cent for capacitance, and 0.3 per cent for dissipation factor.

A test specimen was selected which would reflect the configuration of an electrical apparatus; this is shown in Figure 1. The specimen consists of a metal spool of $1^3/_{16}$ inches diameter wound with about 120 turns of number 26 Formex-insulated wire. Its capacitance to ground is approximately 1,000 micromicrofarads.

Time Variation of Dielectric Properties

Utilizing this experimental technique, the time variations in several of the elec-

---

Paper 53-125, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted October 21, 1952; made available for printing December 10, 1952.

EMANUEL L. BRANCATO is with the Naval Research Laboratory, Washington, D. C.

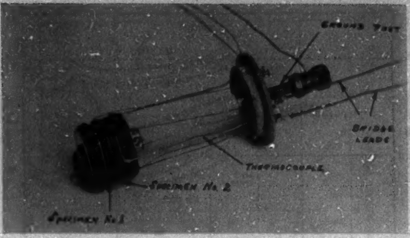
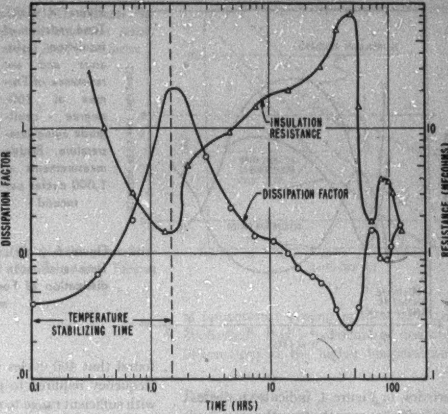


Figure 1 (above). Test specimen

Figure 3 (right). Time variations in dissipation and insulation resistance of Formex at 220-degree-centigrade aging temperature. Bridge measurements at 1,000 cycles per second



trical properties of the dielectric were obtained. Figure 2 is representative of the effects of time and temperature on capacitance. It will be observed that the capacitance rises as the coil is being heated. This rise reflects the bulk expansion of the coil which tends to increase the area and decrease the dielectric spacing. Later in time the plasticizer begins to evaporate and aging effects become evident. As a result of shrinkage during this latter period, capacitance begins to fall with time. Although there is no regular plasticizer in Formex wire enamel, there are high boiling point ingredients, for example, cresol which acts like a plasticizer, and following volatilization a loss of weight by the insulant can be measured. For brevity these ingredients are referred to as plasticizers.

Tracing the effects of temperature and time on dielectric loss, it is observed in Figure 2 that the dissipation factor increases sharply as the specimen temperature rises. Before stabilized conditions are reached, the loss factor curve effectively indicates a temperature dependence starting with room temperature and assuming aging effects are negligible. Increased losses are due, in part, to further decreases in viscosity of the plasticizer which allows greater freedom to the permanent dipolar molecules to move in the presence of an alternating field. As the temperature rises further it will be noticed (at the point where a decrease in capacitance is observed) that the dissipation begins to diminish as a consequence of the inability of the dipoles to follow the field due to thermal agitation resulting in a decrease of frictional losses. The decline in this loss function at these high temperatures is modified by the motion of ions which contribute additional losses. After temperature stabilization at 180 degrees centigrade the loss factor curve is effected by aging time. Essentially, there is a decrease in the loss factor maximum (high temperature absorption peak) with a shift of the peak to higher temperatures, while the temperature of the minimum loss factor point rises indicating that the material remains in the embrittled state at high temperature.[9] At a constant aging temperature, loss of plasticizer continues which increases the viscosity of the dielectric restricting the vibration of the dipoles and ion migration. The decrease in dielectric loss continues with time and reaches a minimum at about 270 hours. This minimum in dissipation factor denotes embrittlement. The loss curve begins to rise again due to decomposition of the material.

The variation in dissipation, during an aging cycle, has been ascribed to variations in the mobility of dipoles and ions. It is conceivable, therefore, that a measure of d-c losses, essentially due to ion migration, should reveal similar behavior. Continuous measurements of d-c resistance was obtained on Formex specimens concurrently with capacitance and dissipation measurements. The results indicated that insulation resistance varied with aging time as an inverse function of the dissipation. This is shown in Figure 3.

Having established the behavior of the dielectric loss, due to ion migration, in Formex, it becomes of interest to obtain a comparative measure of a-c and d-c losses with aging time. These losses are presented as equivalent parallel resistances; the a-c resistance is obtained by calculating a relation incorporating the values of capacitance and dissipation obtained from bridge measurements while the d-c resistance represents the insulation resistance measured during the aging cycles. Examination of the two resistance charac-

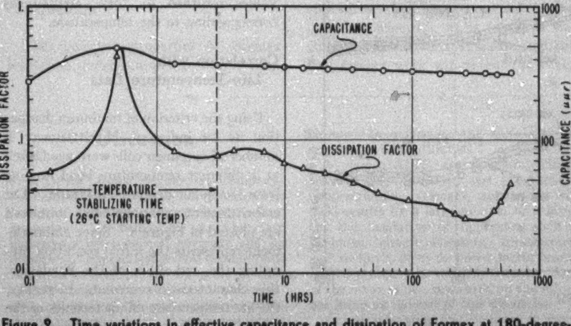


Figure 2. Time variations in effective capacitance and dissipation of Formex at 180-degree-centigrade aging temperature. Bridge measurements at 1,000 cycles per second

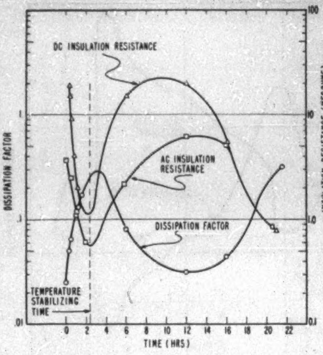


Figure 4 (left). Time variation of insulation resistance and a-c resistance of Formex at 260-degree-centigrade aging temperature. Bridge measurements at 1,000 cycles per second

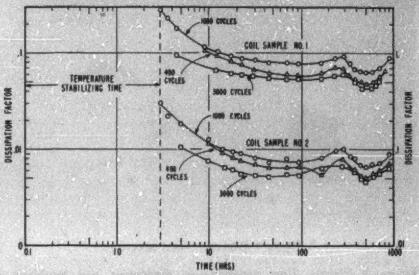


Figure 6 (right). Time variations in dissipation of Formex at 200-degree-centigrade aging temperature. Bridge measurements at 400, 1,000, and 3,000 cycles per second

teristics, in Figure 4, indicates a distinct similarity between the two, the magnitude of the a-c effective parallel resistance naturally being the smaller. At a point beyond embrittlement, the insulation resistance characteristics merge with the a-c resistance characteristics suggesting a disappearance of dipolar losses leaving the losses resulting essentially from ion migration.

Effect of Frequency on Dielectric Measurements

Measurements were made to discover the effect of frequency on the variations of the dielectric characteristics with age. While the capacitance range of the bridge is unaffected by frequency, the dissipation factor range is directly proportional to frequency. Thus, the dissipation factor range is 3 per cent at 50 cycles and 560 per cent at 10 kilocycles. It was found that 400 cycles was the minimum frequency required to provide the bridge with sufficient range to measure the loss in Formex insulation. As the frequency was increased, the bridge was found to be insensitive beyond 3,000 cycles.

Employing this range, data are presented in Figures 5 and 6 of the variations in capacitance and dissipation of two specimens of several, simultaneously aged at 200 degrees centigrade. Comparison of data reveals no distinct shifts in characteristics with frequency suggesting independency of frequency within this frequency spectrum. Furthermore, the corresponding characteristics of the two samples are, for practical purposes, superimposable; thus, one sample is sufficient to furnish representative data. All subsequent tests are conducted at 1,000 cycles.

In view of the limited frequency range of the present instrumentation, no attempt was made to measure relaxation time.

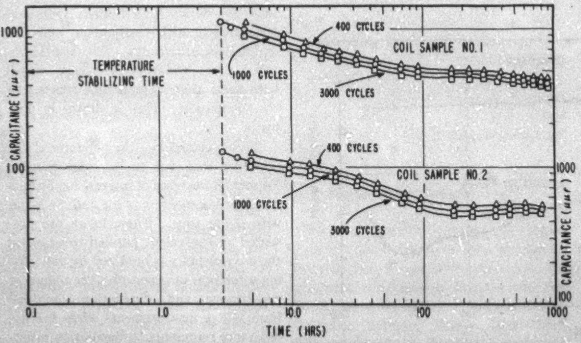


Figure 5. Time variations in the effective capacitance of Formex at 200-degree-centigrade aging temperature. Bridge measurements at 400, 1,000, and 3,000 cycles per second

Effect of Humidity on Dielectric Measurements

Insulation resistance measurements reflect conduction through insulation, as well as conduction over the surface, due to humidity. Correspondingly, dissipation is a measure of all the losses in the material, as well as that loss due to adsorbed humidity. The validity of this is established with the presentation of Figure 7. This figure represents graphically the variations in insulation resistance and dissipation with temperature and humidity. The coil was cycled in temperatures from 30 to 160 degrees centigrade and the measurements of the parameters obtained at suitable intervals. Upon completion of this cycle the specimen was immersed in water for a period of 1/2 hour. Upon removal, the coil was heated and the same measurements were taken. Below 100 degree centigrade no measurements could be obtained because the insulation resistance was very low (the bridge cannot be balanced), but beyond 100 degrees the insulation resistance and dissipation recuperated and the values returned to "dry" magnitudes corresponding to the temperature.

Correlation with Life-Temperature Data

Using the criterion of minimum dissipation as an index of embrittlement, a number of specimen coils were aged, each at a different temperature level ranging from 180 to 260 degrees centigrade. The embrittlement characteristics obtained are plotted in Figure 8. Since embrittlement represents the initial condition preceding the end-of-life of the insulation, this characteristic represents, in reality, the life temperature characteristics of the insulation. This conclusion is supported by the comparison of the life tempera-

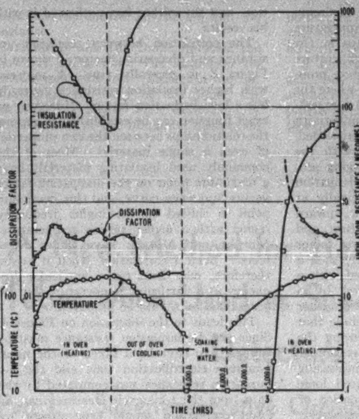


Figure 7 (left). Effects of humidity on insulation resistance and dissipation factor

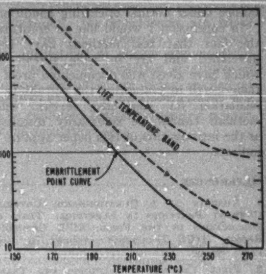


Figure 8 (right). Embrittlement characteristics of Formex

ture characteristics, also in Figure 8, obtained for the same size and type of wire by an earlier investigation.[10] The aging band represents the first and last failure of 20 specimen coils subjected to the stresses experienced by a continuously operating electric motor.

Comparison of the embrittlement curve to the minimum life curve of the aging data reveals that the former is similar in trend but lower in magnitude. This difference is expected because dielectric dissipation is a sensitive function of embrittlement and registers the effect before the cracking of insulation sheathing becomes apparent. Percentagewise, the time elapsed between embrittlement and flaking is greatest at the higher temperature, due to an apparent elasticity resulting from high temperature operation which keeps the insulation or its disintegrated products in literally molten state.

Conclusions

The dielectric properties of Formex insulation undergo a change during thermal aging. This is particularly evident in the dielectric loss and in the insulation resistance. The variations in these values during aging tests can be correlated with the mechanical properties of the materials making it possible to establish the time at which the plasticizer has been removed and embrittlement exists. This point of embrittlement correlates with minimum point on the normal life temperature curves for Formex insulation.

The usefulness of these changes in dielectric properties during thermal aging in determining the true picture of life remaining in insulation can be determined only through further studies of other insulations with dipolar as well as those with nonpolar groups. An investigation of the effects of environment and conditions of contamination would be desirable.

It is important that studies be made of the change in properties of insulated structures which are composed of two or more dielectrics, as is found in machinery construction. It is expected, too, that further work will reveal the effects of temperature cycling on these dielectric properties; this is particularly desirable information at the higher temperatures.

References

1. Field Testing of Generator Insulation, AIEE Committee Report. *AIEE Transactions* (*Electrical Engineering*), volume 60, December 1941, pages 1003–09.

2. Effect of Temperature on Deterioration of Fibrous Insulating Materials, T. Kujirai, T. Akahua. *Bulletin*, Institute of Physics and Chemical Research (Tokyo, Japan), volume 2, 1924–25.

3. The Dielectric Properties of Insulating Material, E. J. Murphy, S. O. Worgan. *Bell System Technical Journal* (New York, N. Y.), volumes 16 and 17, 1937–38.

4. A Relation Between the Power Factor and the Temperature Coefficient of the Dielectric Constant of Solid Dielectrics, W. Gevers. *Phillips Research Reports* (London, England), volume 1, articles I through V, October 1945.

5. Polar Molecules, P. Debye. The Chemical Catalog Company, Inc., 1929.

6. Liquid Dielectrics. *Transactions*, The Faraday Society (London, England), April 1946.

7. Report on Dielectrics, *Transactions*, The Faraday Society (London, England), April 1946.

8. The Basis for the Non-destructive Testing of Insulation, R. F. Field. *AIEE Transactions* (*Electrical Engineering*), volume 60, September 1941, pages 890–95.

9. The Thermal Evaluation of Formvar Resin, T. D. Callinan. *Journal*, Electrochemical Society (New York, N. Y.), volume 98, 1951, pages 171–176.

10. Aging Characteristics of Electrical Insulation, R. E. Whipple. *Report 3708*, Naval Research Laboratory (Washington, D. C.), July 24, 1950.

◆

Discussion

F. S. Oliver (Doble Engineering Company, Belmont, Mass.): The author presents the results of some laboratory tests on Formex insulation. The title of his paper however indicates a more general coverage and he dismisses as impractical several test methods which have been proved by years of trial and experience to produce valuable information concerning the suitability for continued service of electrical power apparatus insulation. Moreover the author does not appear to present a test program which can be used to determine the suitability for service of such insulation.

A recent AIEE questionnaire[1] clearly showed that insulation test programs at nondestructive voltages are used with good results by a large number of utilities. On the insulation of apparatus such as bushings, circuit breakers, transformers, and so forth, 60-cycle power-factor testing has been used with eminent success. Some of the many papers presented on the subject are listed at the end of this discussion.[2–6]

In any insulation test program it should be borne in mind that it is most important to be able to locate the presence of contaminants such as moisture or the results of mechanical injury both of which may be serious failure hazards.

It appears evident that the author carried out his tests on small samples in the laboratory with the ultimate destruction of the sample. I would like to ask the author if the techniques described could be practically applied to nondestructively testing the insulation of all types of power apparatus.

I would also like to ask what tests the author used to determine when embrittlement occurred and when such embrittlement was sufficient to cause insulation failure under normal operating conditions.

In conclusion I would like to emphasize the fact that test methods have been developed and are in wide spread use which have given well documented proof of their worth in evaluating the suitability of insulation for continued service. Such methods cannot be summarily dismissed as the introduction of this paper appears to indicate.

REFERENCES

1. Results of a Questionnaire Covering Current Practices in Electrical Tests on Dielectrics in the Field, **AIEE Committee Report.** *AIEE Transactions*, volume 70, 1951, pages 986–89.

2. Field Testing of Bushings and Transformer Insulation by the Power-Factor Method, **I. W. Gross.** *AIEE Transactions* (*Electrical Engineering*), volume 57, 1938 (October Section), pages 589–600.

3. Field Power-Factor Testing of Transformer Insulation and Operating Experience **E. W. Whitmer.** *AIEE Transactions*, volume 60, 1941, pages 605–11.

4. Bushing and Associated Insulation Testing by the Power-Factor Method, **C. C. Baltzly, E. L. Schlottere.** *AIEE Transactions* (*Electrical Engineering*), volume 60, June 1941, pages 308–12.

5. The A-C Dielectric-Loss and Power-Factor Method for Field Investigation of Electrical Insulation, **Frank C. Doble.** *AIEE Transactions* (*Electrical Engineering*,) volume 60, October 1941, pages 934–39.

T. W. Dakin (Westinghouse Electric Corporation, East Pittsburgh, Pa.): Mr. Brancato has demonstrated in his paper a means of indicating the aging of a particular wire enamel in a coil system. This method may also in some cases be able to predict the failure. Its generality or limitation in this latter respect needs to be tested more extensively since, if it does prove more generally applicable, it merits considerable attention on the part of engineers who must maintain the insulation in such apparatus. It would serve as a monitoring method for inspecting the degree of aging in service of motors, magnet coils, and so forth. A general decline in dissipation factor of motor insulation with aging at elevated temperatures based on a study of two systems of varnished impregnated enamel wire motor insulation has been reported. This decline was similar to the declining dissipation factor observed by Mr. Brancato. I do think, however, that previous investigators have also noted in some cases declining dissipation factor with aging of insulation, although the reason for the declining dissipation factor may have varied from the present case.

Several effects coincident with aging in different varnishes will produce a decreasing ionic conductivity. One of these effects is the slow loss of solvent, or of plasticizer as mentioned by the author. This increases the internal viscosity. Oxidative reactions which crosslink and further polymerize the varnish also will increase the internal viscosity. All of these lead to a decreased ionic conductivity and dissipation factor or increased resistance.

The increased internal viscosity also affects the position of the dissipation factor maximum due to dipolar rotation in the varnish by shifting this maximum toward higher temperatures. An examination of the data published by T. D. Callinan[1] on the dissipation factor of formvar resin would seem to indicate that this maximum occurs at a considerably lower temperature than was used by Mr. Brancato in the present aging studies. After temperature stabilization, his measurements were probably made on a part of the curve where the power factor was increasing with temperature due to increasing ionic conductivity with temperature. As the aging proceeded, the ionic conduction decreased with the hardening of the resin. As the aging proceeds still further, the dipolar dissipation factor maximum which was originally at lower temperatures would shift upward toward the measuring temperature and cause an increasing trend in power factor leading to a minimum. However, such a minimum would not be associated with a maximum d-c resistance followed by a sharply declining value such as the author found. Therefore, we must conclude that the rising dissipation factor following the minimum value and the declining resistance is due not to dipoles but to an increasing ionic conduction, possibly increased ionic concentration, as a result of oxidation or decomposition. It is not readily apparent, however, why such a minimum should necessarily be associated with embrittlement, unless at this point cracking has already started to occur and conduction starts to occur within and on the surface of the cracks. A microscopic examination of the wire just following this point might reveal the cracks. The discusser feels that more evidence and study should be made before there is justification for concluding that such a dissipation factor minimum as the author observed is unequivocally associated with embrittlement of the resin.

In Figure 8, the author has plotted the time to reach the "embrittlement point" and the range of life on a logarithmic scale versus temperature on a linear scale. If the same data are plotted with the temperature on a reciprocal absolute temperature scale, it becomes almost linear and is thereby easier to extrapolate. This indicates that the aging follows a single chemical reaction rate mechanism.

REFERENCE

1. See reference 9 of the paper.

B. P. Kang (Anaconda Wire and Cable Company, Hastings-on-Hudson, N. Y.): First of all, the author must be complimented for his accomplishments in this timely and important investigation. The experimental results are what should be normally expected from both theoretical standpoint and past experience. A few remarks may be in order in regard to the explanations of some of the phenomena.

On Figure 2, the shape of the dissipation factor versus aging time curve is quite a common one. The rise on dissipation factor from 0.1 to 0.5 hour is due to the rise of temperature within this period of time. From 0.5 to about 350 hours, the curve declines first rapidly and then slowly. This seems due to the driving off of the volatile constituents of the insulation; therefore the viscosity is increased. The final rise in dissipation factor from 350 to 600 hours is due to the thermal deterioration of the material. This is a well-known phenomenon and a well-accepted explanation, and the author has confirmed it with his results.

The correlation between insulation resistance and dissipation factor as shown in Figure 3 is generally true. A material with higher insulation resistance generally has a low dissipation factor. However, no exact function can be established to express the relationships between the two properties of even a single material. Most of the commonly used insulating materials have a maximum point on the dissipation factor versus frequency curve, and this maximum point is shifted to the higher frequency region with a higher testing temperature. Murphy and Morgan have shown these facts.[1] Such a correlation, when it exists, therefore, may hold for a particular frequency at a particular temperature, but no generalization should be proposed.

This leads to the discussion on Figure 4. Since the d-c insulation resistance must be supposed to have been taken at the usual 1-minute electrification time and the a-c insulation resistance was computed by the data from the 1-kilocycle measurement, no exact correlation should exist except by coincidence. In order to make a dependable comparison, oscillographic observation on d-c and Schering Bridge measurements on a-c must be taken simultaneously, and the Whitehead-Bannos analysis must be applied. It can be assured that by using the method developed by J. B. Whitehead, all the losses from an a-c measurement can be accounted for.

REFERENCE

1. See reference 3 of the paper.

Emanuel L. Brancato: In closing the discussion, the author wishes to express his appreciation of the interest shown in the subject and of the comments presented by Dr. Dakin, Dr. Kang, and Mr. Oliver.

Unfortunately, the introduction to this investigation has justifiably disturbed Mr. Oliver in that it appeared to condemn all existing nondestructive testing as impractical. This, of course, was not intended. A careful reading of the paper would reveal the purpose of the study to be directed towards the development of a technique capable of predicting expected life of insulation in contradistinction to the objective of existing nondestructive testing which is to determine the present condition of insulation. To offer an analogy, the dielectric measurements have been currently used as a physician uses a thermometer to determine whether or not the patient is fit to work. Continuing with this analogy, the dielectric measurements, in this paper, are used in the same manner an insurance man employs a calendar to predict the expected life.

Dr. Dakin presents an interesting interpretation of the mechanism of the dielectric changes during an aging period of insulation. In the continuation of the experimental work further study will be made to determine the respective contribution of the dipolar and ionic motion to the variation of over-all loss with time.

The author concurs with Dr. Kang's observation relative to the absence of exact relationship between insulation resistance and dissipation factor and is grateful for the offered suggestion on a method of simultaneously measuring the a-c and d-c resistance.

Mr. Oliver and Dr. Dakin question the degree of aging when the insulation reflected a minimum loss. Though it was not indicated in the body of this paper, the insulation specimens were examined for embrittlement at the minimum dissipation point. In all cases a gentle bending of the wire resulted in flaking of insulation demonstrating the embrittlement of the resin.

A question has been raised as to the generality of the reported findings and their applicability to insulation structures employed in existing electrical apparatus. At this point of the study no attempt is made to generalize; the conclusions are based entirely on the properties of Formex. Other insulations of dipolar and nonpolar nature as well as insulation structures of composite dielectric are being studied at present to determine if a similar correlation will be exhibited between dielectric properties and insulation age.

A New Method for Treating Electron Tubes when Used as Superregenerative Detectors

Part I. Superregenerative Circuits Under No-Signal Condition

ABD EL-SAMIE MOSTAFA
NONMEMBER AIEE

M. EL-SHISHINI
NONMEMBER AIEE

Synopsis: This paper is the first of a series of papers dealing with a comprehensive study of superregenerative detection. It deals with superregenerative circuits under no-signal condition; the second paper[1] deals with the same circuits under signal condition, and the third paper[2] is an experimental investigation and verification for the theoretical analysis of papers 1 and 2.

A new theoretical analysis based on treating the electron tube as a periodically varying element is given herein. The methods of superposition and successive approximation are applied to the solution of circuits having periodically varying resistances and the results are directly applied to the present nonlinear problem.

Multiple resonance, the characteristic noise, the building up of oscillation (to an amplitude depending fundamentally on the quenching source and the nonlinearity of the tube characteristic), and the condition of stability of superregeneration are clearly shown. In superregeneration, both the amplitude and frequency of oscillation are modulated; however, the amount of frequency modulation is small. The degree of modulation depends not only on the amplitude and frequency of the quenching voltage, but also on the nonlinearity of the tube characteristic as well as on other circuit parameters. Numerical examples are given to clarify the theoretical predictions.

SUPERREGENERATIVE receiving circuits have become of increasing importance due to their extension into many war-time applications. Accordingly, a need is felt for a rigorous treatment of these circuits.

Several prewar and after-war papers[3-7] dealing, theoretically and experimentally, with superregenerative circuits have been published. In all of these publications, the circuits were treated as if they were linear, and therefore a clear and complete explanation could not be given regarding most superregeneration phenomena.

Theoretical Analysis

As a considerable improvement in performance can be obtained by using a separate tube to produce the required quenching voltage, the following analysis will be confined to a separately quenched superregenerative detector.

A schematic diagram for a separately quenched superregenerative detector is shown, see Figure 1. The losses in the circuit are represented by a shunt arm of conductance G in order to simplify the analysis, and the results obtained apply as well to the more usual case where the losses are associated with the inductance. A sinusoidal quenching voltage $v_q(=N\times \cos pt)$ is applied in the plate circuit of the detector; if it were applied in the grid circuit, it could be very approximately transferred to the plate circuit with its amplitude only multiplied with the amplification factor μ. The automatic bias which is developed under working conditions is represented by a constant bias E_C plus a periodic part of fundamental periodicity equal to the frequency of the quench source and which can be transferred to the plate circuit and then included in v_q. This implies that with a constant quench signal, v_q may vary with different oscillating conditions, variable M for example. This also applies to E_C. Applying Kirchhoff's first law for the currents meeting at the point P we get

$$i_C+i_G+i_p+i_L=0$$

giving after differentiation,

$$d^2v/dt^2+(1/C)\{G\,dv/dt+di_p/dt\}+\omega_0^2v=0 \quad (1)$$

where i_C, i_L, and i_G represent the alternating components of the currents in the capacitive inductive and conductive branches respectively. i_p represents the deviation of the plate current from its steady value, which latter corresponds only to the polarizing potential $(E_b-\mu E_C)$ (plate side), and therefore i_p includes the variation due to the quenching voltage v_q. v is the voltage across the tuned circuit, and $\omega_0^2=1/LC$. i_p can be represented as a power series of the lumped voltage ϵ' as follows

$$i_p=\alpha\epsilon'+\beta\epsilon'^2+\gamma\epsilon'^3 *$$

where

$$\epsilon'=(1-\mu M/L)v+v_q$$

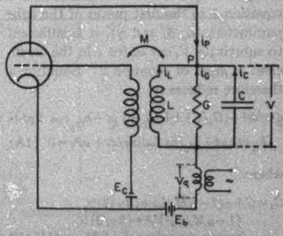


Figure 1. Separately quenched superregenerative detector

---

* For very small amplitudes, $i_p\approx\alpha\epsilon'=\epsilon'/\rho$ and $\alpha=1/\rho$, where ρ = plate resistance of tube.

Paper 53-158, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted March 1, 1950; made available for printing December 3, 1952.

Abd El-Samie Mostafa is with Alexandria University, Alexandria, Egypt. M. El-Shishini is a member of the Egyptian Senate, Cairo, Egypt.

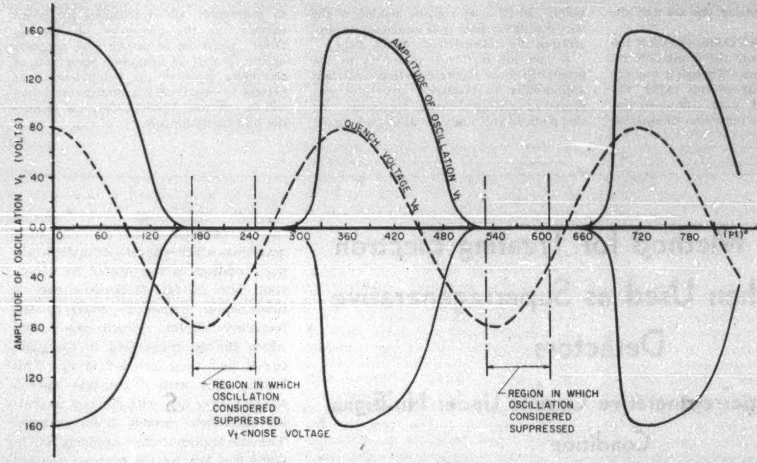


Figure 2. Envelope of oscillation, quench frequency = 100 kilocycles per second

giving

$$dip/d\epsilon' = \alpha + 2\beta\{(1-\mu M/L)v + v_q\} + 3\gamma\{(1-\mu M/L)^2 v^2 + 2(1-\mu M/L)vv_q + v_q^2\}$$

and

$$d\epsilon'/dt = (1-\mu M/L)dv/dt + dv_q/dt$$

If the frequency $p/2\pi$ of the quench source is much less than $\omega/2\pi$ which is usually the case in superregeneration, dv_q/dt can be neglected compared with dv/dt, and

$$d\epsilon'/dt \approx (1-\mu M/L)dv/dt$$

giving

$$\begin{aligned} dip/dt = (dip/d\epsilon')d\epsilon'/dt = \{&\alpha(1-\mu M/L) + 2\beta(1-\mu M/L)^2 v + 2\beta(1-\mu M/L)v_q + \\ &3\gamma(1-\mu M/L)^3 v^2 + 6\gamma(1-\mu M/L)^2 v\, v_q + 3\gamma(1-\mu M/L)v_q^2\}dv/dt \end{aligned} \quad (2)$$

The output voltage v can be assumed to be of the form

$$V_t \cos \omega t + V_{t2} \cos (2\omega t + \theta_{t2}) + V_{t3} \cos (3\omega t + \theta_{t3}) + \ldots$$

where V_t, V_{t2}, V_{t3}, θ_{t2}, θ_{t3}, and ω are slowly varying with time. For the solution of equation 1 to the first power of the tube parameters (α, β, and γ), it is sufficient to substitute $V_t \cos \omega t$ for v in the coefficient of dv/dt of equation 2. Equation 1 therefore reduces to

$$d^2v/dt^2 + (G_t/C)(1 + h_{p_t} \cos pt + h_{2p_t} \cos 2pt + h_{1_t} \cos \omega t + h_{2_t} \cos 2\omega t)dv/dt + \omega_0^2 v = 0 \quad (1A)$$

where

$$G_t = G + \alpha(1-\mu M/L) + (3\gamma/2)\times(1-\mu M/L)^3 V_t^2 + (3\gamma/2)\times(1-\mu M/L)N^2$$

$$h_{p_t} = (1/G_t)2\beta(1-\mu M/L)N$$

$$h_{2p_t} = (1/G_t)3\gamma(1-\mu M/L)(N^2/2)$$

$$h_{1t} = (1/G_t)(2\beta + 6\gamma N \cos pt)(1-\mu M/L)^2 V_t$$

and

$$h_{2t} = (1/G_t)3\gamma(1-\mu M/L)^3(V_t^2/2).$$

In this way, the nonlinear differential equation 1 has been transformed to a linear one with periodically varying coefficients.

In the present analysis, it is assumed that each of $(G_t/\omega C)$, $(G_t/\omega C)h_{1_t}$, and $(G_t/\omega C)h_{2_t}$ is much less than unity (this is equivalent to saying that the present work is limited to small signals).

For the solution of equation 1(A), the methods of superposition and successive approximation will be adopted. Although these methods can be applied directly to equation 1(A), a misunderstanding may result due to the fact that the amplitudes of variation h_{1_t}, h_{2_t} and so forth are functions of the required amplitude of oscillation V_t. Moreover, the coefficient of v in equation 1(A), namely, "ω_0^2" is constant which is not the case under signal condition; therefore the general case where such a coefficient is variable with time should be treated. Again, as a series circuit is usually preferred to its dual parallel one, the former will be chosen for the solution of equation 1(A). Consider a series $L-C_t-R_t$ circuit, where R_t is a function of time and given by

$$R_t = R_{0t}\left\{1 + \sum_{m=1}^{\infty} h_{mp_t} \cos mpt + \sum_{n=1}^{\infty} h_{nt} \cos n\omega t\right\}$$

R_{0t}, h_{mp_t}, and h_{n_t} are slowly varying with time; C_t is also a slow function of time and the inductance L is constant, and $p << \omega$. The differential equation for the charge Q across capacitor C_t is given by

$$d^2Q/dt^2 + (R_{0t}/L)\left(1 + \sum_{m=1}^{\infty} h_{mp_t} \cos mpt + \sum_{n=1}^{\infty} h_{nt} \cos n\omega t\right) dQ/dt + \omega_{0t}^2 Q = 0 \quad (3)$$

where

$$\omega_{0_t}^2 = 1/LC_t$$

Although successive approximation can be applied to equation 3, the amount of work can be reduced by the transformation $Q = q\epsilon^F$, where F is a function of time. Substitution in equation 3 gives

$$d^2q/dt^2 + \{2dF/dt + R_t/L\}dq/dt + \{d^2F/dt^2 + (dF/dt)^2 + (R_t/L)dF/dt + \omega_{0t}^2\}q = 0 \quad (4)$$

F is so chosen that the coefficient of dq/dt in equation 4 vanishes; therefore

$$F = -\int (R_t/2L)dt = -\int (R_{0t}/2L)\times\left\{1 + \sum_{m=1}^{\infty} h_{mp_t} \cos mpt + \sum_{n=1}^{\infty} h_{nt} \cos n\omega t\right\} dt$$

Coefficient of q in equation 4 is given by

$$\omega_{0t}^2\left[1 - (1/2\omega_{0t}^2 L)(d/dt)\left\{\sum_{m=1}^{\infty} R_{0t}h_{mp_t}\times \cos mpt + \sum_{n=1}^{\infty} R_{0t}h_{nt} \cos n\omega t\right\} - (R_{0t}/2\omega_{0t}L)^2\left\{1 + \sum_{m=1}^{\infty} h_{mp_t} \cos mpt + \sum_{n=1}^{\infty} h_{n_t} \cos n\omega t\right\}^2\right]$$

Now, if $(R_{0t}/2\omega_{0t}L)$ is much less than unity and ω is near ω_{0t}, both assumptions being satisfied in the present nonlinear problem, terms containing $(R_{0t}/2\omega_{0t}L)^2$ can be neglected. Again, as R_{0t}, h_{nt}, ω, and h_{mp_t} are very slow functions of time, that is, their variation in one cycle of the frequency $\omega/2\pi$ can be neglected, and also as $p \ll \omega$, coefficient of q will be given approximately by

$$\omega_t^2\left\{1+(R_{0t}/2\omega_tL)(\omega/\omega_t)\sum_{n=1}^{\infty} nh_{nt}\sin n\omega t\right\}$$

where

$$\omega_t^2=\omega_{0t}^2-(1/2L)(d/dt)\sum_{m=1}^{\infty} R_{0t}h_{mp_t}\cos mpt$$

Equation 4, therefore, becomes

$$diq/dt+\omega_t^2\left\{1+(R_{0t}/2\omega_tL)(\omega/\omega_t)\sum_{n=1}^{\infty}\times nh_{nt}\sin n\omega t\right\}\int iqdt=0 \quad (5)$$

where $iq=dq/dt$, and where R_{0t}, ω, ω_t, and h_{nt} are very slow functions of time; therefore, for a small interval of time t_1, corresponding to several cycles of oscillations, they can be considered constants and equal R_0, ω, ω_0, and h_n respectively. Equation 5 then becomes

$$diq/dt+\omega_0^2\left\{1+(R_0/2\omega_0L)(\omega/\omega_0)\sum_{n=1}^{\infty}\times nh_n\sin n\omega t\right\}\int iqdt=0 \quad (5A)$$

The current iq of 5(A) to a good degree of approximation, so long as each of $(h_1R_0/\omega L)$, $(h_2R_0/\omega L)$, and so forth is much less than unity, will be given by (see Appendix I):

$$iq=B_1\epsilon^{\sqrt{a}t}[-(R_0/2\omega L)h_1\cos(\phi'-\theta')/\times 2\{(\sqrt{a}/\omega)+\omega/\sqrt{a}\}\sqrt{1+a/\omega^2}+\sin(\omega t+\phi')-\{\sqrt{M_2^2+N_2^2}\omega^2\times\sqrt{4+a/\omega^2}/2\sqrt{a^2+9\omega^4+10a\omega^2}\}\times\sin(2\omega t+\theta_2+\theta_2'-\psi_2)-\{\sqrt{M_3^2+N_3^2}\omega^2\sqrt{9+a/\omega^2}/2\times\sqrt{a^2+64\omega^4+20a\omega^2}\}\sin(3\omega t+\theta_3+\theta_3'-\psi_3)-\ldots] \quad (6)$$

where

$$\sqrt{a}=(\omega/4)\sqrt{(h_2R_0/\omega L)^2-4(2d+d^2)^2},\quad \tan\phi'=2(2d+d^2)/\{(4\sqrt{a}/\omega)+(h_2R_0/\omega L)\},\ \tan\theta'=\omega/\sqrt{a}$$

$$M_2=(1/\sqrt{1+a/\omega^2})(h_1+3h_3)(R_0/2\omega L)\times\sin(\phi'-\theta')$$

$$N_2=-(1/\sqrt{1+a/\omega^2})(h_1-3h_3)\times(R_0/2\omega L)\cos(\phi'-\theta')$$

$$\tan\theta_2=N_2/M_2,\quad \tan\theta_2'=2\omega/\sqrt{a},\quad \tan\psi_2=4\omega\sqrt{a}/(a-3\omega^2)$$

$$M_3=(1/\sqrt{1+a/\omega^2})(h_2+2h_4)\times(R_0/\omega L)\sin(\phi'-\theta')$$

$$N_3=-(1/\sqrt{1+a/\omega^2})(h_2-2h_4)\times(R_0/\omega L)\cos(\phi'-\theta')$$

$$\tan\theta_3=N_3/M_3$$

$$\tan\theta_3'=3\omega/\sqrt{a}$$

and

$$\tan\psi_3=6\omega\sqrt{a}/(a-8\omega^2)$$

The charge

$$q=\int i_qdt=(B_1/\omega)\epsilon^{\sqrt{a}t}[-\{\cos(\phi'-\theta')/2\times(1+a/\omega^2)^{3/2}\}(h_1R_0/2\omega L)+\sin(\omega t+\phi'-\theta')/\sqrt{1+a/\omega^2}-\{\omega^2\sqrt{M_2^2+N_2^2}/2\times\sqrt{a^2+9\omega^4+10a\omega^2}\}\sin(2\omega t+\theta_2-\psi_2)-\{\omega^2\sqrt{M_3^2+N_3^2}/2\times\sqrt{a^2+64\omega^4+20a\omega^2}\}\sin(3\omega t+\theta_3-\psi_3)-\ldots] \quad (6A)$$

The charge q given by equation 6(A) is true for the small time interval t_1 which contains several cycles of oscillations. For a general solution at any time t, $\sqrt{a}t$ should be replaced by $\int\sqrt{a}\,dt$ and ωt by $\int\omega dt$. Charge q will be given by

$$q=(B_1/\omega)\epsilon^{\int\sqrt{a}dt}[-\{\cos(\phi'-\theta')/2\times(1+a/\omega^2)^{3/2}\}(h_{1t}R_{0t}/2\omega L)+\sin(\int\omega dt+\phi'-\theta')/\sqrt{1+a/\omega^2}-\{\omega^2\sqrt{M_2^2+N_2^2}/2\times\sqrt{a^2+9\omega^4+10a\omega^2}\}\sin(\int 2\omega dt+\theta_2-\psi_2)-\{\omega^2\sqrt{M_3^2+N_3^2}/2\times\sqrt{a^2+64\omega^4+20a\omega^2}\}\sin(\int 3\omega dt+\theta_3-\psi_3)-\ldots]$$

The charge Q of equation 3 will, therefore, be given by

$$Q=\epsilon q^{\int[\sqrt{a}-(R_{0t}/2L)\{1+\sum_{m=1}^{\infty}h_{mp_t}\times\cos mpt+\sum_{n=1}^{\infty}h_{n_t}\cos n\omega t\}]dt}$$
$$\approx(B_1/\omega)\epsilon^{\int(\sqrt{a}-R_{0t}/2L)dt}[-h_{1t}(R_{0t}/\times 2\omega L)\cos(\phi'-\theta')+\sin(\int\omega dt+\phi'-\theta')+(h_{1t}R_{0t}/6\omega L)\cos(\int 2\omega dt+\phi'-\theta')+(h_{2t}R_{0t}/16\omega L)\times\cos(\int 3\omega dt+\phi'-\theta')+\ldots] \quad (7)$$

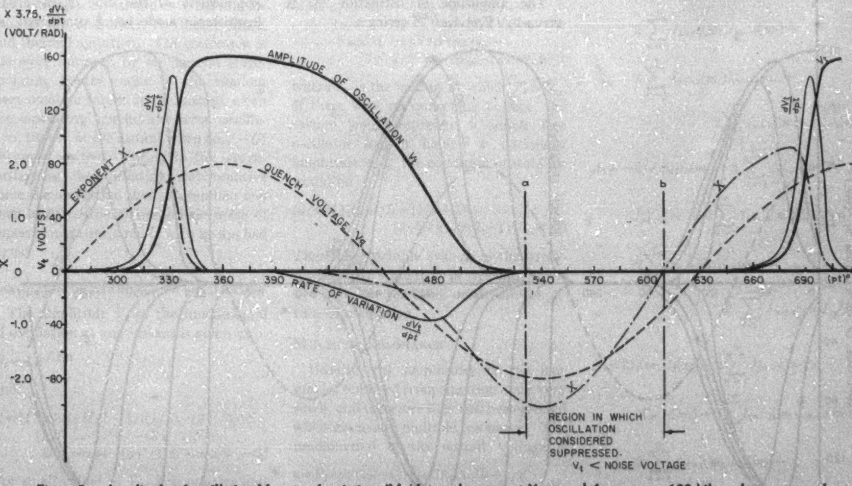


Figure 3. Amplitude of oscillation V_t, rate of variation dV_t/dpt, and exponent X, quench frequency = 100 kilocycles per second

V_t (at a) = V_t (at b) = noise voltage

The amount of detune d is obtained from the relation $\omega_t=\omega(1+d)$ where

$$\omega_t=\omega_{0t}\times \sqrt{1-(1/2\omega_{0t}^2L)(d/dt)\sum_{m=1}^{\infty}R_{0t}h_{m_{p_t}}\cos mpt}$$

$$\approx\omega_{0t}\left\{1-(1/4\omega_{0t}^2L)(d/dt)\sum_{m=1}^{\infty}\times R_{0t}h_{m_{p_t}}\cos mpt\right\}$$

and

$$\omega_{0t}^2=1/LC_t$$

Now, equation 3 is similar to 1(A) with the substitution of Q, R_{0t}, L, and C_t for v, G_t, C, and L respectively. Also the assumptions made for the solution of equation 3 are satisfied in equation 1(A). However, the voltage v must be of the same form as assumed and which is $v=V_t\cos\omega t+\ldots.$ Therefore, a possible solution of 1(A) can be obtained from equation 7 provided $\phi'-\theta'=\pi/2$, giving $d\approx-(1/8)(G_t/\omega C)^2h_{2t}^2\approx0$, and hence

$$\omega\approx\omega_t=\omega_0\left\{1-(1/4)(1/\omega_0^2L)(d/dt)\sum_{m=1}^{\infty}\times R_{0t}h_{m_{p_t}}\cos mpt\right\}$$

where

$$\omega_0^2=1/LC$$

Therefore, the output voltage v of equation 1(A) is given by

$$v=A_s\epsilon^{\int X dt}\{\cos(\omega_0t+\Delta\omega)-(1/6\omega C)\times(2\beta-6\gamma'N\cos pt)(1-\mu M/L)^2\times V_t\sin 2(\omega_0t+\Delta\omega)-(\mu M/L-1)^3\times(1/8\omega C)(3\gamma'V_t^2/4)\times\sin 3(\omega_0t+\Delta\omega)-\ldots\}\quad(8)$$

where

$V_t=A_s\epsilon^{\int X dt}$

A_s = a constant depending on initial conditions = amplitude of oscillation at start

$$X=(1/2C)[\alpha'-G-(3\gamma'/4)(\mu M/L-1)^3\times V_t^2-(3\gamma'/2)(\mu M/L-1)N^2+(\mu M/L-1)\{2\beta N\cos pt-(3\gamma'/2)N^2\cos 2pt\}]$$

$$\Delta\omega=-\{(\mu M/L-1)/4\omega_0C\}\{-2\beta N\cos pt+(3\gamma'/2)N^2\cos 2pt\}$$

$\alpha'=\alpha(\mu M/L-1)$, and $\gamma'=-\gamma$

It should be noted that the system is, in general, oscillating in the absence of the quenching voltage and hence $\mu M/L>1.0$. Also, for all practical cases, γ is negative and β is positive. It is clear from equation 8 that both the amplitude and frequency of oscillation are modulated. The fundamental frequency of modulation equals the quenching voltage frequency. The amount of modulation depends not only on the amplitude and frequency of the quenching voltage, but also on the nonlinearity of the tube characteristic as well as other circuit parameters.

ENVELOPE OF OSCILLATION V_t AND THE RATE OF VARIATION OF THE AMPLITUDE OF OSCILLATION dV_t/dpt

The amplitude of oscillation V_t is given by, $V_t=A_s\epsilon^{\int X dt}$ giving

$$dV_t/dpt=XV_t/p=(\epsilon/2)(\omega/p)\times[1-(V_t^2/V_0^2)-\{3\gamma'(\mu M/L-1)/\times 2(\alpha'-G)\}N^2(1+\cos 2pt)+\{2\beta(\mu M/L-1)/(\alpha'-G)\}N\cos pt]V_t\quad(9)$$

where

$$X=(1/2C)\{\alpha_{eq}'-G-(3\gamma'/4)(\mu M/L-1)^3V_t^2\}$$

$$\epsilon=(\alpha'-G)/\omega C$$

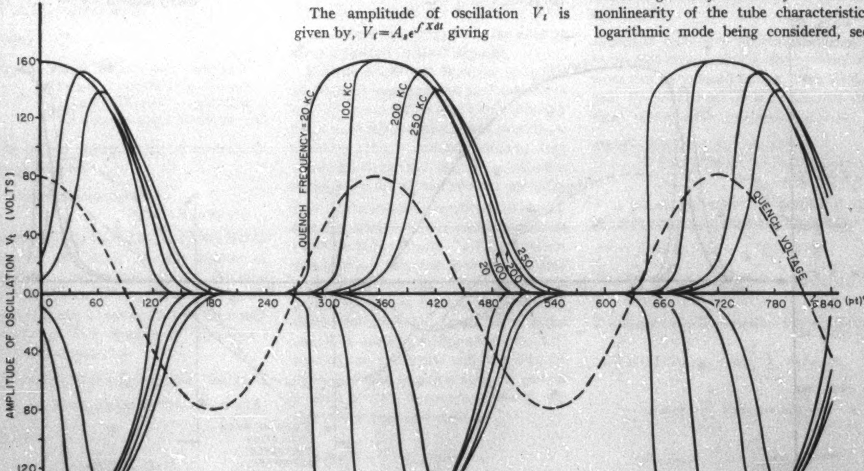
and as the amount of nonlinearity is assumed small, $\epsilon<<1$

$$\alpha_{eq}'=(\mu M/L-1)\{(1/\rho)-(3\gamma'/2)N^2+2\beta N\cos pt-(3\gamma'/2)N^2\cos 2pt\}$$

and $V_0=\sqrt{(\alpha'-G)/(3/4)\gamma'(\mu M/L-1)^3}$ is the amplitude of free oscillation if the quench voltage were absent. Comparing X with the corresponding exponent of a free oscillator,[8] it is concluded that a superregenerative detector can be considered as an ordinary oscillator with periodically variable regeneration. The exponent X depends not only on the amplitude N and the frequency $p/2\pi$ of the quench source, but also on the nonlinearity of the tube characteristic and the amplitude of oscillation, Figure 3.

In superregeneration, the amplitude N and the frequency $p/2\pi$ of the quench source are in general so adjusted that the exponent X during a quench cycle decreases, reaches zero, and reverses its sign, see Figure 3. During a part of the quench cycle, X is positive and oscillation builds up exponentially to an amplitude which is generally limited by N and the nonlinearity of the tube characteristic, logarithmic mode being considered, see

Figure 4. Envelope of oscillation at different quench frequencies



Figures 2 and 3. At the other part of the quench cycle, X is negative and oscillation dies down to an amplitude of the order of the internal tube and external circuit noise voltages, at which interval the oscillation can be considered suppressed, see Figures 2 and 3. During the interval at which $(\alpha_{eq}'-G)$ is positive, the system works as an oscillator with variable regeneration. When $(\alpha_{eq}'-G)$ is negative, the oscillation will be damped to practically zero amplitude and remaining practically suppressed till $(\alpha_{eq}'-G)$ becomes positive, and the system starts oscillating again. The same process is repeated for each cycle of quench voltage.

It is clear from equation 9 that the effect of increasing the quench frequency $p/2\pi$ is to reduce the rate of variation dV_t/dpt. If $p/2\pi$ is increased beyond a certain value, the amplitude of oscillation cannot attain the saturation value and the detector will operate in the linear mode instead of the logarithmic mode as shown in Figure 4 (characteristics with $p/2\pi$ equal to 200 and 250 kilocycles per second).

THE CHARACTERISTIC NOISE

It should be noted that if there were not any electromotive force impressed upon the circuit, the oscillation could not start even if $(\alpha_{eq}'-G)$ became positive. But when no signal is applied to the detector, a number of random electromotive forces act upon the oscillatory circuit from various internal and external sources such as the shot effect, the flicker effect, and thermal agitation. The oscillation is therefore initiated by an impulse corresponding to the value of the random electromotive forces at the instant when the operating conditions favor oscillation, that is, at the instant when $(\alpha_{eq}'-G)$ becomes positive. Owing to the irregularities in the initiating electromotive force the oscillation starts at random and thus leads to the characteristic noise of superregenerative circuits with no applied signal.

CONDITION OF STABILITY OF OSCILLATION

The amplitude V_t of the fundamental of oscillation at any instant is given by

$$V_t=A_s\epsilon^{\int X dt}$$

and

$$X=(1/2C)[(\mu M/L-1)\{(1/\rho)-(3\gamma'/4)\times (\mu M/L-1)^2V_t^2-(3\gamma'/2)N^2+ 2\beta N\cos pt-(3\gamma'/2)N^2\cos 2pt\}-G]$$

For small values of V_t, X is positive and the oscillation builds up from the small starting value A_s which corresponds to the noise voltage. As time goes on, $\int X dt$ increases and hence V_t increases; however, at the same time, X reduces. Therefore $\int X dt$ tends towards a maximum value, at which instant V_t becomes maximum and equals V_m. Therefore, V_t is maximum when $\int X dt$ is maximum and equals V_m/A_s, that is, when $X=0$. Alternatively, V_t is maximum when $dV_t/dt=0$, but $dV_t/dt=XV_t$; thus, V_t is maximum when $X=0$. Maximum values of V_t can therefore be obtained from the following relation

$$(X)_{V_t=V_m}=(1/2C)\{(\mu M/L-1)/\rho-G- (3\gamma'/4)V_m^2(\mu M/L-1)^3-(\mu M/ L-1)\{(3\gamma'/2)N^2-2\beta N\cos pt+ (3\gamma'/2)N^2\cos 2pt\}]=0 \quad (10)$$

Equation 10 as it is, gives one value for V_m, this is because the tube characteristic was represented by three terms only for simplicity. But, if more terms were taken, equation 10 would yield more than one value for V_m. Whether these maximum values occur in practice or not depends on their stability.

If there is a slight increase δV in the possible maximum amplitudes V_m given by equation 10, then the increase δX in $(X)_{V_t=V_m}$ must be negative, that is, X must be reduced from its zero value to a negative one, otherwise $\epsilon^{\int X dt}$ would increase beyond V_m/A_s and hence the maximum amplitude would rise indefinitely which is contrary to the fact. Therefore the condition for stability is, $(dX/dV_m)<0$, that is

$$(d/dV_m)\{(\mu M/L-1)/\rho-G-(3\gamma'/4)\times (\mu M/L-1)^3V_m^2\}<-(d/dV_m)\times \{(\mu M/L-1)(2\beta N\cos pt_m- (3\gamma'/2)N^2\cos 2pt_m-(3\gamma'/2)N^2)\}$$

where t_m is the instant at which $V_t=V_m$. If there were no quenching voltage, the system would represent a simple free oscillator and if having a stationary amplitude $=V_m$, the condition of stability would be

$$(d/dV_m)\{(\mu M/L-1)/\rho-G- (3\gamma'/4)(\mu M/L-1)^3V_m^2\}<0$$

Therefore, a stable system in the superregenerative state may be unstable in the free state for the same maximum amplitude of oscillation.

MULTIPLE RESONANCE

Usually, the amplitudes of the harmonics obtained from equation 8 are very small, and therefore they can be neglected when discussing multiple resonance. The fundamental of the output voltage is

$$v=V_t\cos\{\omega_0 t-(1/4\omega_0 C)(\mu M/L-1)\times (-2\beta N\cos pt+(3\gamma'/2)N^2\cos 2pt)\}$$

The amplitude of oscillation V_t is given by

$$V_t=A_s\epsilon^{\int X dt}$$
$$=A_s\epsilon^{\int (1/2)\{m_0+pm_1\cos pt+2pm_2\cos 2pt-(3\gamma'/4C)(\mu M/L-1)^2V_t^2\}}$$

giving

$$(d/dt)(1/V_t^2)+(m_0+pm_1\cos pt+2pm_2\times \cos 2pt)(1/V_t^2)=(3\gamma'/4C)(\mu M/L-1)^2$$

and

$$1/V_t^2=\epsilon^{-(m_0t+m_1\sin pt+m_2\sin 2pt)}\times \{B+(3\gamma'/4C)(\mu M/L-1)^2\times \int\epsilon^{(m_0t+m_1\sin pt+m_2\sin 2pt)}dt\}$$
$$=B\epsilon^{-m_0t}\{A_0+A_1\cos pt+A_2\cos 2pt+ \ldots -B_1\sin pt-B_2\sin 2pt-\ldots\}+ (3\gamma'/4C)(\mu M/L-1)^2[(A_0/m_0)+ \{(m_0A_1-pB_1)/(p^2+m_0^2)\}\cos pt+ \{(m_0A_2-2pB_2)/(4p^2+m_0^2)\}\cos 2pt+ \ldots +\{(m_0B_1+pA_1)/(p^2+m_0^2)\}\times \sin pt+\{(m_0B_2+2pA_2)/(4p^2+m_0^2)\}\times \sin 2pt+\ldots]\{A_0+A_1\cos pt+ A_2\cos 2pt+\ldots -B_1\sin pt- B_2\sin 2pt-\ldots\} \quad (11)$$

This gives the transient and the steady-state solutions respectively where $B=$a constant depending on initial conditions.

$$m_0=(1/C)\{\alpha'-G-(3\gamma'/2)(\mu M/L-1)N^2\}$$

$$m_1=(2\beta/pC)(\mu M/L-1)N$$

$$m_2=-(3\gamma'/4pC)(\mu M/L-1)N^2$$

$$A_0=I_0(m_1)I_0(m_2)-2\sum_{n=1}^{\infty}I_{8n-4}(m_1)I_{4n-2}(m_2)+ 2\sum_{n=1}^{\infty}I_{8n}(m_1)I_{4n}(m_2)$$

$$A_1=2\sum_{n=1}^{\infty}I_{8n-7}(m_1)I_{4n-3}(m_2)- 2\sum_{n=1}^{\infty}I_{8n-5}(m_1)I_{4n-3}(m_2)- 2\sum_{n=1}^{\infty}I_{8n-3}(m_1)I_{4n-1}(m_2)+ 2\sum_{n=1}^{\infty}I_{8n-1}(m_1)\ I_{4n-1}(m_2)$$

$$A_2=-2I_2(m_1)I_0(m_2)+2\sum_{n=1}^{\infty}I_{8n-6}(m_1)\times I_{4n-2}(m_2)+2\sum_{n=1}^{\infty}I_{8n-2}(m_1)I_{4n-2}(m_2)- 2\sum_{n=1}^{\infty}I_{8n-2}(m_1)I_{4n}(m_2)- 2\sum_{n=1}^{\infty}I_{8n+2}(m_1)I_{4n}(m_2)$$

$$B_1=2I_1(m_1)I_0(m_2)-2\sum_{n=1}^{\infty}I_{8n-5}(m_1)\times I_{4n-2}(m_2)-2\sum_{n=1}^{\infty}I_{8n-3}(m_1)I_{4n-1}(m_2)+ 2\sum_{n=1}^{\infty}I_{8n-1}(m_1)I_{4n}(m_2)+ 2\sum_{n=1}^{\infty}I_{8n+1}(m_1)I_{4n}(m_2)$$

$$B_2=2\sum_{n=1}^{\infty} I_{8n-8}(m_1)I_{4n-3}(m_2)-$$
$$2\sum_{n=1}^{\infty} I_{8n-4}(m_1)I_{4n-3}(m_2)-$$
$$2\sum_{n=1}^{\infty} I_{8n-4}(m_1)I_{4n-1}(m_2)+$$
$$2\sum_{n=1}^{\infty} I_{8n}(m_1)I_{4n-1}(m_2)$$

and I_n = modified Bessel function of order n.[9] As stated before, in some parts of the quench cycle, the oscillation can be considered suppressed, in which part the amplitude of oscillation is less than the internal tube noise and external circuit noise voltages. Therefore, in each quench cycle, the oscillation is started by an amplitude due to the noise voltage, and therefore both the steady and transient solutions exist.

Now, if steady state only is considered

$$v=(A_0'+A_1'\cos pt+A_2'\cos 2pt+\ldots+ B_1'\sin pt+B_2'\sin 2pt+\ldots)\times \cos\{\omega_0t-((\mu M/L-1)/4\omega_0C)\times (-2\beta N\cos pt+(3\gamma'/2)N^2\cos 2pt)\}$$

where A_0', A_1', $A_2'\ldots$ and B_1', $B_2'\ldots$ are functions of m_0, p, A_1, $A_2\ldots$, B_1, $B_2\ldots$ Therefore

$$v=(A_0'+A_1'\cos pt+A_2'\cos 2pt+\ldots+ B_1'\sin pt+B_2'\sin 2pt+\ldots)\times$$
$$\Bigg[\!\Bigg[(\cos\omega_0t)\Bigg[\Bigg\{J_0(m_1p/4\omega)+(-1)^n\times \sum_{n=1}^{\infty}J_{2n}(m_1p/4\omega)\cos 2npt\Bigg\}\times$$
$$\Bigg\{J_0(m_2p/4\omega)+(-1)^n2\sum_{n=1}^{\infty}\times J_{2n}(m_2p/4\omega)\cos 4npt\Bigg\}-$$
$$\Bigg\{(-1)^{n+1}2\sum_{n=1}^{\infty}J_{2n-1}(m_1p/4\omega)\cos npt\Bigg\}\times$$
$$\Bigg\{(-1)^{n+1}2\sum_{n=1}^{\infty}J_{2n-1}(m_2p/4\omega)\times \cos npt\Bigg\}\Bigg]+(\sin\omega_0t)\Bigg[\Bigg\{(-1)^{n+1}2\times$$
$$\sum_{n=1}^{\infty}J_{2n-1}(m_1p/4\omega)\cos npt\Bigg\}\times$$
$$\Bigg\{J_0(m_2p/4\omega)+(-1)^n2\sum_{n=1}^{\infty}\times J_{2n}(m_2p/4\omega)\cos 4npt\Bigg\}+$$
$$\Bigg\{J_0(m_1p/4\omega)+(-1)^n2\sum_{n=1}^{\infty}\times J_{2n}(m_1p/4\omega)\cos 2npt\Bigg\}\Bigg\{(-1)^{n+1}2\times$$
$$\sum_{n=1}^{\infty}J_{2n-1}(m_2p/4\omega)\cos 2npt\Bigg]\Bigg\}\Bigg]\!\Bigg]$$

where J_n is a Bessel function of order n.[9]

But, as p/ω is $<<1$, therefore terms containing $J_n(mp/4\omega)$ can be neglected and the term $J_0(mp/4\omega)$ can be considered unity. Therefore, the voltage v, to a good degree of approximation, may be given by

$$\begin{aligned}v&\approx(A_0'+A_1'\cos pt+A_2'\cos 2pt+\ldots+ B_1'\sin pt+B_2'\sin 2pt+\ldots)\cos\omega_0t\\ &=A_0'\cos\omega_0t+(A_1'/2)\{\cos(\omega_0+p)t+ \cos(\omega_0-p)t\}+(A_2'/2)\{\cos(\omega_0+2p)t+ \cos(\omega_0-2p)t\}+\ldots+(B_1'/2)\times\\ &\quad\{\sin(\omega_0+p)t-\sin(\omega_0-p)t\}+ (B_2'/2)\{\sin(\omega_0+2p)t- \sin(\omega_0-2p)t\}+\ldots\end{aligned}$$

This shows that the resonance curve of the circuit represents a major resonance at $\omega_0/2\pi$, plus resonances of diminishing importance at $(\omega_0\pm np)/2\pi$, clarifying that multiple resonance is intrinsically a part of superregeneration. It is clear that such multiple resonance is mainly due to amplitude modulation.

Numerical Examples

In order to clarify the foregoing theoretical predictions, several examples will be calculated. Let

$\omega/2\pi=75$ megacycles per second

$\epsilon=(\alpha'-G)/\omega C=0.01$

$V_0=100$ volts

$\{3\gamma'(\mu M/L-1)/2(\alpha'-G)\}N^2=0.5$

$\{2\beta(\mu M/L-1)/(\alpha'-G)\}N=2.5$

Substitution in equation 9 gives

$$dV_t/dpt=\{3.75\times 10^5/(p/2\pi)\}\{0.5(1+ 5\cos pt-\cos 2pt)-V_t^2/10^4\}V_t \quad (12)$$

Although there is a complete solution given by equation 11 for the determination of V_t, yet it is easier to obtain V_t graphically from equation 12 using the step-by-step method. As stated before, both the transient and steady states exist and therefore the solution depends on the initial conditions, namely, the amplitude due to the impulse of the noise voltage at the instant at which the exponent X starts to be positive and which will be assumed to be 0.1 volt.

For a quench frequency of 100 kilocycles per second, the envelope of oscillation is shown in Figure 2, the quench voltage is also shown. The oscillation builds up from an amplitude = 0.1 volt (assumed noise voltage) at the instant of the quench cycle corresponding to $(250°+m\times 360°)$, that is at the instant at which the exponent X starts to be positive; the oscillation dies down to the same amplitude at the instant corresponding to $(170°+m\times 360°)$, where m is an integer. In the interval $(170°+m\times 360°)$ up to $(250°+m\times 360°)$, the oscillation is considered suppressed. Figure 3 shows the amplitude of oscillation V_t, the rate of amplitude variation dV_t/dpt and the exponent X, quench frequency being 100 kilocycles per second. Figure 4 shows the effect of the quench frequency on the envelope of oscillation. It is clear that at the quench frequencies 200 and 300 kc per second, the detector operates in the linear mode. At 20 and 100 kc per second, it operates in the logarithmic mode.

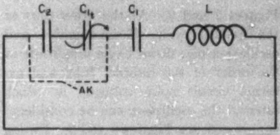


Figure 5. Connection diagram for equation 5(A)

Appendix I. Determination of the Current *iq* of Equation 5(A)

Let $\omega_0=\omega(1+d)$ where d = amount of detune (a small fraction). The connection diagram giving the current iq of equation 5(A) will be as shown, see Figure 5, where

$1/C_1=\omega^2L$

$1/C_2=(2d+d^2)/C_1$

and

$$1/C_{1t}=(1/C_1)(R_0/2\omega L)\sum_{n=1}^{\infty}nh_n\sin n\omega t$$

According to the well-known compensation theorem, if C_2 and C_{1t} are short-circuited by closing the artificial key (a.k.) and replaced instead by a generator of zero internal impedance and having an electromotive force $-\{(1/C_2)+(1/C_{1t})\}\int iq\,dt$ where iq is the actual current flowing in the circuit, iq would be unchanged; but the actual current flowing is unknown. Successive approximations will be adopted as follows.

Closing the a.k., the current would be given by, $i_1=A\sin(\omega t+\phi)$, where A and ϕ are constants to be determined from initial conditions; the condenser is assumed to be initially charged. i_1 gives the first approximation for the current iq.

For a second approximation, replace C_2 and C_{1t} by a generator of zero internal impedance and having an electromotive force

$$\begin{aligned}v_1&=-\{(1/C_2)+(1/C_{1t})\}\int i_1dt\\ &=A\{(2d+d^2)/\omega C_1\}\cos(\omega t+\phi)+ (A/2\omega C_1)(R_0/2\omega L)\times\\ &\quad[h_1\{\sin(2\omega t+\phi)-\sin\phi\}+ 2h_2\{\sin(3\omega t+\phi)+\sin(\omega t-\phi)\}+ 3h_3\{\sin(4\omega t+\phi)+\sin(2\omega t-\phi)\}+ \ldots]=v_{10}+v_{1\omega}+v_{2\omega}+v_{3\omega}+\ldots\end{aligned}$$

The voltage v_1 consists of components of frequencies zero (that is, d-c component), $\omega/2\pi$, $2\omega/2\pi$, $3\omega/2\pi$, and so forth.

It should be noted that the voltage component $v_{1\omega}$ is of the same frequency as the natural frequency of the $L-C_1$ circuit. Therefore, if this component were applied to the $L-C_1$ circuit, the resulting current would increase indefinitely. In fact, it is this component which has the most important effect in producing current.

To get the effect of $v_{1\omega}$, close the a.k. of Figure 5 and apply such a voltage, other generators being short-circuited, the current is given by

$$i_2=-(A/2)\omega^2[2(2d+d^2)\{-(t/2\omega)\times \cos(\omega t+\phi)+(1/4\omega^2)\sin(\omega t+\phi)\}-(R_0/\omega L)h_2\{(t/2\omega)\sin(\omega t-\phi)+(1/4\omega^2)\cos(\omega t-\phi)\}]$$

The components of i_2 whose amplitudes do not contain the time t can be included in $i_1(=A\sin(\omega t+\phi))$ in which case A and ϕ will be different, but yet depend on initial conditions. Therefore, the components whose amplitudes contain the time t will only be considered. The effects of the components of v_1 of frequencies other than $\omega/2\pi$ on the current fundamental can be neglected so long as each of $(R_0/\omega L)h_1$, $(R_0/\omega L)h_2$, $(R_0/\omega L)h_3$, and so forth is much less than unity; conditions which are satisfied in the present nonlinear problem.

Therefore, the fundamental component of the current iq to a second degree of approximation may be given by, $i_1+i_2=A\times\{\sin(\omega t+\phi)+(2(2d+d^2)/4)\omega t\cos(\omega t+\phi)+(R_0/\omega L)h_2(\omega/4)t\sin(\omega t-\phi)\}$. Following the same lines, other approximations can be obtained and which may be given by

$$i_3=A[\{(R_0/\omega L)h_2(\omega/4)\}^2-\{2(2d+d^2)/4\}^2\omega^2](t^2/2!)\sin(\omega t+\phi)$$

$$i_4=A[\{(R_0/\omega L)h_2(\omega/4)\}^2-\{2(2d+d^2)\omega/4\}^2](t^3/3!)\times\{(2(2d+d^2)\omega/4)\cos(\omega t+\phi)+(R_0/\omega L)h_2(\omega/4)\sin(\omega t-\phi)\}$$

$$i_5=A[\{(R_0/\omega L)h_2(\omega/4)\}^2-\{2(2d+d^2)\omega/4\}^2]^2(t^4/4!)\times\sin(\omega t+\phi)\ldots$$

Therefore, if each of $2(2d+d^2)\omega/4$, $(R_0/\omega L)h_1$, $(R_0/\omega L)h_2$, and so forth is very small compared with unity, the fundamental of the current iq, to a satisfactory degree of approximation, is given by

$$\begin{aligned}iq_{\mathrm{Fund.}}&=i_1+i_2+i_3+i_4+i_5+\ldots\\&=A\{1+(\sqrt{a})^2(t^2/2!)+(\sqrt{a})^4(t^4/4!)+\ldots\}\sin(\omega t+\phi)+A\{2(2d+d^2)\omega/4\}\times(1/\sqrt{a})\{\sqrt{a}t+(\sqrt{a})^3(t^3/3!)+\ldots\}\cos(\omega t+\phi)+A\{(R_0/\omega L)\times(h_2\omega/4)(1/\sqrt{a})\{\sqrt{a}t+(\sqrt{a})^3\times(t^3/3!)+\ldots\}\sin(\omega t-\phi)\end{aligned}$$

$$iq_{\mathrm{Fund.}}=A[\cosh(\sqrt{a}t)\sin(\omega t+\phi)+\{(\sinh(\sqrt{a}t))/\sqrt{a}\}\{(2(2d+d^2)\times\omega/4)\cos(\omega t+\phi)+(R_0/\omega L)\times(h_2\omega/4)\sin(\omega t-\phi)\}]$$

where

$$\sqrt{a}=(\omega/4)\sqrt{(h_2R_0/\omega L)^2-\{2(2d+d^2)\}^2}$$

The current $iq_{\mathrm{Fund.}}$ can be written as

$$iq_{\mathrm{Fund.}}=B_1\epsilon^{\sqrt{a}t}\sin(\omega t+\phi')+B_2\epsilon^{-\sqrt{a}t}\times\sin(\omega t+\phi'')$$

where

$$\tan\phi'=2(2d+d^2)/\{(4\sqrt{a}/\omega)+(R_0h_2/\omega L)\}$$

and

$$\tan\phi''=-2(2d+d^2)/\{(4\sqrt{a}/\omega)-(h_2R_0/\omega L)\}$$

The angles ϕ' and ϕ'' are independent of initial conditions, they depend only on the amplitude of variation h_2 and the amount of detune d. However, B_1 and B_2 can be considered as constants depending on initial conditions. As $\sqrt{a}$ is real, which is the case of interest at present, the term containing $\epsilon^{-\sqrt{a}t}$ can be neglected, as it is a highly damped one, and $iq_{\mathrm{Fund.}}$ can be satisfactorily given by

$$iq_{\mathrm{Fund.}}=B_1\epsilon^{\sqrt{a}t}\sin(\omega t+\phi')$$

The harmonics other than the fundamental as well as the zero-frequency component can be obtained as follows.

Replace C_2 and C_{1t} by a generator of zero internal impedance and having an electromotive force equal to

$$\begin{aligned}&-\{(1/C_2)+(1/C_{1t})\}\int iq_{\mathrm{Fund.}}dt\\&=-\{B_1\epsilon^{\sqrt{a}t}/\sqrt{1+a/\omega^2}\}[\{2(2d+d^2)/2\omega C_1\}\sin(\omega t+\phi'-\theta')-(h_1R_0/2\omega L)(1/2\omega C_1)\{\cos(2\omega t+\phi'-\theta')-\cos(\phi'-\theta')\}-(2h_2R_0/2\omega L)\times(1/2\omega C_1)\{\cos(3\omega t+\phi'-\theta')-\cos(\omega t-\phi'+\theta')\}-(3h_3R_0/2\omega L)\times(1/2\omega C_1)\{\cos(4\omega t+\phi'-\theta')-\cos(2\omega t-\phi'+\theta')\}-\ldots]\\&=v_{0h}+v_{1h}+v_{2h}+v_{3h}+\ldots\end{aligned}$$

where $\tan\theta'=\omega/\sqrt{a}$.

This voltage consists of components of frequencies zero, $\omega/2\pi$, $2\omega/2\pi$, $3\omega/2\pi$, and so forth, being equal to v_{0h}, v_{1h}, v_{2h}, v_{3h}, and so forth, respectively.

The effect of the component of frequency $\omega/2\pi$ is included in the predetermined fundamental.

The effects of components of frequencies zero, $2\omega/2\pi$, $3\omega/2\pi$, and so forth, give the zero-frequency, second, third, and so forth harmonic current components respectively to a good degree of approximation so long as each of $h_1R_0/\omega L$, $h_2R_0/\omega L$, $h_3R_0/\omega L$, and so forth is much less than unity. Therefore, the current iq will be given by equation 6 in the text.

References

1. A New Method for Treating Electron Tubes When Used as Superregenerative Detectors. Part II—Superregenerative Circuits Under Signal Condition, **Abd El-Samie Mostafa, M. El-Shishini.** *AIEE Transactions*, volume 52, part I, 1953 (*Paper 53-157*).

2. A New Method for Treating Electron Tubes When Used as Superregenerative Detectors. Part III—Experimental Investigation of Superregenerative Circuits, **Abd El-Samie Mostafa, M. El-Shishini.** *AIEE Transactions*, volume 52, part I, 1953 (*Paper 53-155*).

3. Some Recent Developments on Regenerative Circuits, **E. H. Armstrong.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 10, 1922, page 244.

4. On Superregeneration of an Ultra-Short-Wave Receiver, **H. Ataka.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 23, 1935, page 841.

5. The Basic Principles of Superregenerative Reception, **F. W. Frink.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 26, 1938, page 76.

6. Superregeneration for Broadcast Receivers, **D. Maurice.** *Wireless Engineer* (London, England), volume 15, 1938, page 4.

7. Electronics Manual for Radio Engineers (book). McGraw-Hill Book Company, Inc., New York, N. Y., "Superregenerative Circuit Applications," Harry Stockman, 1949, page 617.

8. A New Method for Treating Electron Tubes when Used as Oscillators in the Free as Well as in the Synchronized States, **Abd El-Samie Mostafa, M. El-Shishini.** *AIEE Transactions*, volume 70, 1951, pages 999–1004.

9. Bessel Functions for Engineers (book), **N. W. McLachlan.** Oxford University Press, Oxford, England, 1946.

No Discussion

Sensitivity and Output Formulas for the Resistance Bridge

PHILIP M. ANDRESS
ASSOCIATE MEMBER AIEE

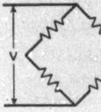
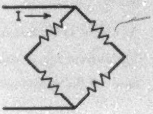
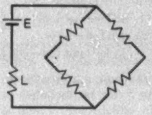
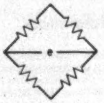
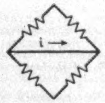
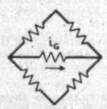
IN EITHER the design or the application of a resistance bridge circuit, it may be necessary to calculate the output of the bridge in terms of input or vice versa. The purpose of the calculation may be to establish the required sensitivity of a galvanometer, meter, or of any other null detector or output device, which is to be energized by a resistance change in the bridge, or it may be necessary to determine the input voltage or current required for minimum operation. Upon occasion, an exact calculation may be required to calibrate the output indicator, as in the deflection types of resistance thermometer or strain-gauge bridges. The formulas and correction factors presented in Table I were developed by the author for convenience in performing such calculations.

Paper 53-1, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted March 25, 1952; made available for printing October 7, 1952.

Philip M. Andress is with the Rubicon Company, Philadelphia, Pa.

The author wishes to express thanks to Dr. Frank Wenner of the National Bureau of Standards and to Norman E. Bonn and other members of the Rubicon Company staff for their invaluable help and suggestions during preparation of this material.

Table I. Sensitivity and Bridge Output Formulas*

Output	Input			Correction Factor for $p=\pm\Delta X/X$†
				
	(1A) $e=\frac{VCp}{T}$	(1B) $e=\frac{IBXp}{T}$‡	(1C) $e=\frac{EBXp}{T(D+L)}$	$1/(1+pw)$
	(2A) $i=\frac{Vp}{T}$	(2B) $i=\frac{IDp}{T}$	(2C) $i=\frac{Ep}{T(1+L/D)}$	$1/(1+py)$
	(3A) $i_G=\frac{Vp}{T(1+G/C)}$	(3B) $i_G=\frac{IBXp}{T(C+G)}$	(3C) $i_G=\frac{Ep}{T(1+G/C)(1+L/D)}$	$1/(1+ps)$

* Error of uncorrected formulas is less than p-proportional parts, hence is negligible for near balance calculations.
† Exact formula = uncorrected formula × correction factor.

Auxiliary Equations:
1. $e=Vp/4$ when $A=B=X=R$ (see item 5, Appendix)
2. $V=ID$
3. $CD=BX=AR$
4. $\sqrt{CD}=\sqrt{BX}=\sqrt{AR}$
5. $e=iC$
6. $w=[X+AD/(D+L)]/T$**
7. $y=[X+R+AD/(D+L)]/T$**
8. $s=[X+RC/(C+G)+AD/(D+L)]/T$**

** If the input is a constant regulated voltage, or if V is the voltage across the unbalanced bridge, then the value of L in the correction factors is zero; if a constant regulated current, or if I is the input current to the unbalanced bridge, then L is infinite and the fractions having $(D+L)$ in the denominator disappear from those expressions. The linear equations 1(C), 2(C), and 3(C) have no significance in these two special cases:

1. If $L=0$, then $w=(X+A)/T$; $y=(X+R+A)/T$; and $s=[X+RC/(C+G)+A]/T$.
2. If $L=\infty$, then $w=X/T$; $y=(X+R)/T$; and $s=[X+RC/(C+G)+A]/T$.

‡ See also item 6, Appendix.

These formulas can be used in essentially three types of applications:

1. The calculation of sensitivity of the balanced or near-balanced bridge in terms of the required precision of measurement or test (examples 1, 2, and 3).

2. The close approximation of the bridge output for moderate unbalances because of resistance changes in one or more arms.

3. The calculation of the exact output for any degree of unbalance because of a change in the X arm (examples 4 and 5).

It will be noted that if the uncorrected formulas are considered only as approximations, the other uncorrected formulas can be derived from equation 1(A) and auxiliary equations 2 and 3. These formulas can also be applied to the double, or Kelvin bridge circuit, except that CD must be substituted for BX where applicable, and T is the sum of the resistances of the six arms consisting of the two double-ratio arms plus the standard and unknown arms. However, they are particularly convenient for the single or Wheatstone bridge and have been developed primarily for application to the latter circuit.

Formula 1(A) is equivalent to the equation given by Wenner.[1] Formulas 3(A) and 3(C) have been previously published.[2] Formula 3(C) with correction factor is the equivalent of the so-called classical solution of the Wheatstone bridge, which because of its complexity is of historical rather than of practical value.[1] Auxiliary equation 1 has been in previous use for some time. Here it is presented as a special case of equation 1(A).

Nomenclature and Definitions Used in Table I

e = available electromotive force output
i = available current output
i_G = current through galvanometer branch (output)
V = input voltage to the balanced bridge
I = input current to the balanced bridge
E = battery electromotive force
C = resistance of the balanced bridge as viewed by the galvanometer (output bridge resistance)
D = resistance of the balanced bridge as viewed by the battery (input bridge resistance)
p = proportional-parts unbalance (see definition following) of the bridge, or the required proportional-parts precision
A, B, R, and X = resistances of the four arms of the balanced bridge, taken in rotation (B and X being opposite arms)
T = total additive resistance of all arms ($T=A+B+R+X$)
L = resistance of the battery branch (or limiting resistance); in general, the resistance looking from bridge toward power source
G = resistance of the galvanometer branch; in general the resistance looking from bridge toward the output
Available electromotive force = open circuit output electromotive force (see item 2, Appendix)
Available current = short-circuit output current (available electromotive force divided by output resistance) (see item 2, Appendix)

Galvanometer branch = galvanometer plus auxiliary series or parallel resistance, if used (output branch)

Battery branch = battery plus limiting resistance (input branch).

Balanced bridge = bridge in which the relationship $A/B = X/R$ holds, hence in which the output is zero, or null. Otherwise the bridge is said to be unbalanced.

Proportional-parts unbalance = the proportional-parts change of resistance in one arm (as $p = \Delta X/X$) to produce the unbalance, or in general the algebraic sum of the proportional parts changes where more than one arm is involved, that is, $p = \Delta X/X - \Delta R/R + \Delta B/B - \Delta A/A$.

Bridge sensitivity = least value of p to produce a detectable output, or alternatively, output for a specified value of p.

Examples

1. A 300-ohm galvanometer is to be used as the null detector for 0.1-per-cent measurements with a Wheatstone bridge. At balance, the resistances of the unknown and rheostat arms are $X = 200{,}000$ ohms and $R = 2{,}000$ ohms. The total resistance of the two ratio arms, $A+B$, is 1,000 ohms. What current will pass through the galvanometer corresponding to an 0.1-per-cent unbalance, if input voltage V is 10 volts, and if the galvanometer is in parallel with $A+B$?

Answer: $V=10$, $p=0.1$ per cent $=0.001$, $T=203{,}000$ ohms, $G=300$ ohms (assuming that no dummy resistance is used), and $C=1{,}000$ ohms. Substituting these values into formula 3(A) gives

$$i_G = (10\times 0.001)/(203{,}000\times 1.3) = 0.04\times 10^{-6} \text{ ampere}$$

2. If a galvanometer having an external critical-damping resistance of 10,000 ohms and an electromotive force sensitivity (see item 3, Appendix) of 100 microvolts per millimeter is to be used across this bridge, together with dummy series resistance to adjust the damping to critical, what deflection will correspond to 0.1-per-cent unbalance?

Answer: From formula 1(A) for available electromotive force

$$e = (10\times 1000\times 0.001)/203{,}000 = 50\times 10^{-6} \text{ volts} = 50 \text{ microvolts}$$

Since this available electromotive force is unaffected by the series resistance, the deflection is 50 microvolts ÷ 100 microvolts/millimeter = 1/2 millimeter.

3. Again referring to the same bridge, but assuming that a galvanometer with an electromotive force sensitivity of 4.5 microvolts per millimeter and an external critical damping resistance of 45 ohms is to be used with dummy parallel resistance for adjustment to critical damping, what deflection will correspond to 0.1-per-cent unbalance?

Answer: From formula 2(A) the available, or short-circuit current is

$$(10\times 0.001)/203{,}000 = 0.05 \text{ microampere}$$

Since this available current is unaffected by parallel resistance, the available electromotive force is $0.05\times 45 = 2.25$ microvolts (auxiliary equation 5). Dividing by the electromotive force sensitivity of 4.5 microvolts per millimeter gives 0.5-millimeter deflection.

4. The output electromotive force of a 120-ohm strain-gauge bridge is measured by means of an accurate potentiometer. Assuming that three arms of the bridge remain constant and that the resistance L in the battery branch can be neglected (see footnote** to Table I), what potentiometer reading will correspond to a 20-per-cent increase in resistance of the strain gauge when the input voltage is 1 volt?

Answer: Use formula 1(A) or auxiliary equation 1 with correction factor for available electromotive force. $V=1$, $p=0.2$. Hence the uncorrected value is $Vp/4 = 0.05$ volt. $w = [X + AD/(D+L)]/T = [120 + 120 \times 120/(120)]/480 = 0.5$. Hence the correction factor is $1/(1+pw) = 1/(1+0.2\times 0.5) = 1/1.1$. The product is the exact value: $0.05\times(1/1.1) = 0.045$ volt.

5. What is the potentiometer reading in this setup for a 20-per-cent decrease in the resistance of the strain gauge?

Answer: Here uncorrected value and the value of w remain unchanged, and p is unchanged except for sign. The correction factor for $p = -0.2$ is $1/(1-0.2\times 0.5) = 1/0.9$; hence the potentiometer reading is $0.05\times(1/0.9) = 0.0555$ volt, with polarity reversed from that of example 4.

The correction factors are also useful to determine the error or nonlinearity of the uncorrected values, again where $p = \Delta X/X$. This error is, as applicable, pw, py, or pz proportional parts of the exact value and since neither w, y, or z can exceed 1.0, the limit of error of any uncorrected equation value where p is small compared to 1.0 is simply p-proportional parts (see item 1, Appendix). To illustrate, in examples 1 through 3, the limit of error is 0.1 per cent. It should be noted that this error represents a change in output corresponding to $(0.1 \text{ per cent})^2$, or to a resistance change of 1 part per million in the X arm, and that the actual error is probably less than this amount. On the other hand, where a calibrated unbalance of more than a few per cent is involved, the nonlinearity of the uncorrected formula may be significant, as in example 4 where the difference between the uncorrected and the exact value is 0.0045 ohm or 9 per cent, and in example 5 is 0.0055 ohm or $11^1/_9$ per cent.

PREFERRED LOCATION OF GALVANOMETER BRANCH IN THE BALANCED BRIDGE

This is a modification of the rule previously presented by Kotter[3] which, contrary to the traditional rule,[5] takes into account both the distribution of wattage in the four arms and the damping of the galvanometer: "Connect the galvanometer across the higher or the lower bridge resistance depending on whether the specified external damping resistance of the galvanometer is higher or lower respectively than the mean bridge resistance."

The mean bridge resistance is defined as the geometric mean between the bridge resistances viewed by the battery and galvanometer branches. Numerically, this is the same as the geometric mean between either pair of opposite arms (auxiliary equation 4). The specified external damping resistance may be taken as the external critical-damping resistance, a published galvanometer characteristic (see item 4, Appendix).

The galvanometer is across the higher bridge resistance when it connects the junction of the two highest resistance arms with the junction of the two lowest resistance arms; otherwise it views the lower bridge resistance.

6. A galvanometer having an external critical damping resistance of 1,000 ohms is to be used at critical damping across a bridge whose arm resistances at balance taken in rotation are 90, 900, 100,000, and 10,000 ohms. Which is the preferred galvanometer connection?

Answer: The mean resistance between a pair of opposite arms is $\sqrt{90\times 100{,}000}$, or 3000 ohms. Since the external critical damping resistance is less than this value,

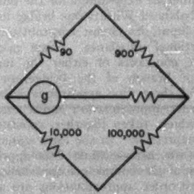


Figure 1

the better connection is across the lower bridge resistance as shown in Figure 1. In this case, the traditional rule would have called for the alternate arrangement, assuming that resistance is higher in the galvanometer branch than in the battery branch.[5]

Appendix

1. In the strictest sense, the limit of error of the uncorrected solution for any value of p, whether large or small and irrespective of sign, is p-proportional parts of the exact solution. However, it may also be considered as p-proportional parts of the uncorrected solution provided the unbalance is small say where p is less than 10 per cent. This over-all limit applies whether the unbalance is due to a resistance change in one arm only, or to a combination of changes in more than one arm.

2. The designations 'available electromotive force' and 'available current' were suggested to me by Dr. Frank Wenner of the National Bureau of Standards.

3. The electromotive force sensitivity of a galvanometer is defined as that electromotive force which produces unit deflection when it is introduced into a series circuit consisting of the galvanometer and its external critical-damping resistance. This electromotive force is also the potential difference appearing across the open circuit produced when the galvanometer is disconnected from this circuit, and is herein designated as the available electromotive force. It is often convenient to consider the electromotive force sensitivity of the galvanometer in the latter sense, as is illustrated by examples 2 and 3.

4. The galvanometer operates at highest efficiency at near-critical damping, hence it should be selected for a match between its external critical-damping resistance and the bridge resistance, rather than between its internal resistance and bridge resistance, as is sometimes recommended. Specifically, efficiency is usually highest when relative damping is between 0.6 and 0.8. Harris[4] gives a detailed presentation of this subject.

5. More generally $e=Vp/4$ when two arms adjacent to the galvanometer branch points are equal even though all four arms are not equal. Another useful equation, which can be derived from equation $2(B)$, is $i=Ip/4$ when two arms adjacent to the battery branch points are equal. When $A=B=X=R$ both $e=Vp/4$ and $i=Ip/4$.

6. It should be noted that formula $1(B)$ may be also conveniently written $e=(Ip)/S$, where S is the sum of the conductances of the four arms, that is, $S=1/A+1/B+1/R+1/X$. The relationship between equation $2(A)$ and this expression of equation $1(B)$ is an interesting illustration of the principal of duality.[6] Equivalent dual relationships apply to the other equations as well.

References

1. Methods, Apparatus, and Procedures for Comparison of Precision Standard Resistors, **Frank Wenner.** *Research Paper RP 1323, Journal of Research* (National Bureau of Standards, Washington, D. C.), volume 25, 1940, equation 10 on page 234, equation 76 on page 277.

2. The Wheatstone Bridge, **Philip M. Andress.** *Notes Number 2,* Rubicon Company (Philadelphia, Pa.), 1951.

3. Location of the Galvanometer Branch for Maximum Sensitivity of the Wheatstone Bridge, **F. Ralph Kotter.** *Research Paper 1844, Journal of Research* (National Bureau of Standards, Washington, D. C.), volume 40, 1948, page 401.

4. Electrical Measurements (book), **F. K. Harris.** John Wiley and Sons, New York, N. Y., 1952, pages 58–60.

5. A Treatise on Electricity and Magnetism, **J. C. Maxwell.** Clarendon Press, Oxford, England, volume I, third edition, 1892, page 478.

6. Elementary Electric-Circuit Theory, **Richard H. Frazier.** McGraw-Hill Book Company, Inc., New York, N. Y., 1945, pages 85–87.

◆

Discussion

F. Wenner and **F. K. Harris** (National Bureau of Standards, Washington, D. C.): The concepts of open-circuit voltage and short-circuit current (referred to in the present paper as available electromotive force and available current respectively) are, of course, well known in the solution of network problems. It has not been nearly so generally recognized that in the solution of the unbalanced Wheatstone bridge they are extremely valuable in that they permit the network and the detector to be considered as separate entities. The simplification that results from this separation and individual consideration of cause and effect is very well illustrated in Mr. Andress' collection of equations, and he is to be congratulated for carrying this work through to its logical conclusion. It should be pointed out that because of space limitations the development of these equations was omitted from the present paper, but we are glad to learn that this development is set forth in detail elsewhere.[1]

In the paper Mr. Andress has tabulated nine different expressions for the output of an unbalanced Wheatstone bridge in terms of its input, its circuit constants, and its proportional unbalance. Some cases also involve constants of either the input or detector branch of the bridge, or both. Each of these nine expressions involves a first-order approximation and a correction term which, taken together, constitute an exact expression. In the form stated, each is new, whether we think of the approximation or of the complete expression. Since the first-order approximations are simpler and serve nearly all needs, we shall limit further consideration to them.

We will recall Mr. Andress' statement of the order of approximation, that the proportional error resulting from use of the approximate expression is no greater than the proportional unbalance of the bridge (that is, no more than a 1-per-cent error results from using the approximate equations if the bridge is out of balance by 1 per cent). Obviously, then, if adjustments are being made to establish balance, the approximate expressions approach exact expressions. Furthermore, this is true even if certain liberties are taken with the expressions: V may represent the potential drop across the bridge whether it is precisely or only approximately balanced; I may represent current through the bridge for an exact or an approximate balance; and C may represent the bridge resistance seen from the galvanometer position for an exact or an approximate balance. Furthermore, it is necessary to know the bridge output in a form that can be used in stating galvanometer sensitivity so that, while it is customary to state galvanometer sensitivity in terms of the relation between its steady deflection and the current in its winding, it is proper and may be more convenient to use the available electromotive force or the available current rather than the current in the winding.

Taking into account the possibility that each of the nine first-order approximations may be stated in other forms, and accepting the modifications stated above for V, I, and C, Andress' equation $1(A)$ has been used by each of us, while equation $3(B)$ has been used by Jaeger. Equations $2(A)$, $2(B)$, and $2(C)$ are definitely new. Presumably equations $1(B)$ and $1(C)$ are also new. Whether $3(A)$ and $3(C)$ are new would be difficult to say, since many writers have derived more or less similar equations. However, what is important is not the novelty of these expressions either in form or in substance but rather their convenience in use. The input, which we will need to know, may be measured with a voltmeter or an ammeter, or we may use simply the rated electromotive force of the battery. The constants of the circuit can usually be calculated from readings of the various bridge dial settings. An additional condition that must be met is that the resistance of the bridge circuit as seen from the position of the galvanometer must be appropriate for proper damping. Furthermore there must be an appropriate limit on the power dissipated in the bridge arms. These additional conditions may be stated explicitly or implicitly.

Andress' nine approximate expressions involve the three ways of expressing input and the three ways of expressing output, and constitute a complete assembly of bridge solutions. There are, of course, many other expressions stating the relations between the input, the circuit constants, and the output of a bridge, but most of these other expressions are more complicated than the simpler of the nine expressions under consideration here. It seems probable that in one way or another equations $1(A)$ and $2(B)$ of the paper will largely replace the more complicated expressions presently available as well as those which have been in general use since before the turn of the century.

REFERENCE

1. I. Deflection Constants of the D-C Galvanometer at a Specified Damping. II. Improved Solutions of the Unbalanced Bridge Circuit, **Philip M. Andress.** *Review of Scientific Instruments* (New York, N. Y.), volume 24, number 2, February 1953, pages 172–77.

Cathode-Ray Synchroscope and Automatic Synchronizer

M. ABDEL-HALIM AHMED
MEMBER AIEE

Synopsis: Two new electronic apparatus are introduced, in which the main element is a cathode-ray tube. The first is a synchroscope. The bus-bar voltage causes the beam to trace a circle, while the incoming machine voltage gives a pulse every cycle. The pulse is either applied to the *Y*-deflection plates or to the grid of the tube. In the first case it appears as a "pip" rotating round the circle, while in the latter case a luminous spot is seen to trace a circle. The behavior of both the pip and the spot is similar to that of the pointer of an ordinary synchroscope. The second apparatus is an automatic synchronizer. It is based on the same principle as the synchroscope, but uses a special cathode-ray tube in which the fluorescent screen is replaced by two collecting electrodes. When the pulse is applied to the grid of the tube, the beam current flows to one electrode or the other, and when it flows to the proper electrode, it causes a relay to operate which in turn causes the synchronizing circuit breaker to close. Both apparatus are discussed, and a description including the design data of an actual experimental synchroscope is given.

IN ALL branches of industry, a modern industrial tool is forging a silent evolution. This tool is electronics. Today there is scarcely a field of industry where electronic equipment is not speeding production, refining the process of manufacture, improving quality, doing efficiently what before had not been done at all, and making possible accurate and reliable measurements where other instruments or apparatus cannot be used or are unreliable. Above all, electronic control equipment has become versatile in all fields. Therefore, the modern tendency in power generating stations is to have the control gear operating electronically whenever possible.

Although the synchroscope is an essential part of a power station, yet synchronizing equipment has not followed the recent rapid developments in science and industry, and electronics has not yet invaded that field.

The cathode-ray synchroscope is the first developed type of electronic synchroscope. It is used in the Communications Laboratories of the Faculty of Engineering at Fouad I University.

The new cathode-ray automatic synchronizer exhibits the main features of an automatic synchronizing equipment, making use of the general principles underlying the cathode-ray synchroscope. It involves a new application of the phenomenon of commutating a cathode-ray beam by means of electrodes replacing the normal screen of a cathode-ray tube.

The Cathode-Ray Synchroscope

General Principle

The idea of this synchroscope lies in having a cathode-ray tube in which the beam traces a circle (or an ellipse) by the use of a conventional phase-splitting resistance-capacitance circuit, giving two voltages with a 90-degree phase difference which are applied to the *X* and *Y* plates of the cathode-ray tube; see Fig. 1. The voltage applied to the phase-splitting circuit is derived from the bus-bar voltage of the power station through an isolating line transformer of a suitable ratio.

The voltage of the incoming line is used to give a pulse at a certain defined instant of its cycle. This pulse is superimposed on the *Y*-plate deflecting voltage obtained from the phase-splitting circuit. Actually two pulses (displaced by 180 degrees) are obtained in each cycle, and therefore the circle traced on the screen of the cathode-ray tube will now have two pips, one positive and the other negative, displaced by 180 degrees, and their position relative to the circle depends on the instantaneous phase difference between the bus bar and incoming line voltages.

When these two voltages have exactly the same frequency, the pips will always occur when the beam is tracing two definite points of the circle, and the trace will be a circle with two stationary pips. If the frequencies of the bus bar and incoming line are different, the two pips will be seen to move round the circle, in the same manner as the pointer of an ordinary synchroscope rotates. The direction of the motion depends on which frequency is greater, and the speed of motion depends on the amount of difference between the two frequencies.

When the two voltages are exactly in phase, the two pips will be at two certain points on the screen; if the two frequencies are equal, the two pips will be stationary at these two points, which indicates the right instant for synchronizing; see Fig. 2(*A*). The two positions of the pips can be adjusted to occur exactly on the *Y* axis of the tube, Fig. 2(*B*), by means of a suitable arrangement of the circuit.

Generation of Pulse

The easiest method for obtaining a pulse, which occurs only twice per cycle and at definite instants in the cycle, is by the use of a saturated transformer, which serves at the same time for the isolation of the cathode-ray tube deflection system from the incoming voltage system.

The sine-wave input voltage is applied to the primary of the transformer through a suitable high resistance, Fig. 3, which ensures that the magnetizing current flowing in the primary winding of the transformer is very nearly sinusoidal. It was found that a transformer with only one or two laminations of good-quality core material gives satisfactory results.

Assuming an ideal magnetization curve for the core material as shown in Fig. 4(*A*), the voltage induced in the secondary winding of the transformer would be as shown in Fig. 4(*B*). The pulses occur at the instants when the primary voltage is zero.

However, due to the effects of the hys-

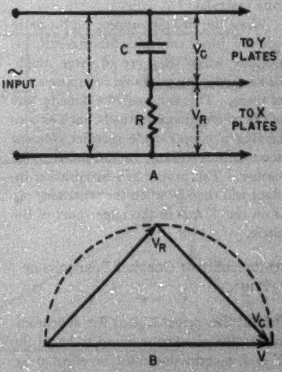
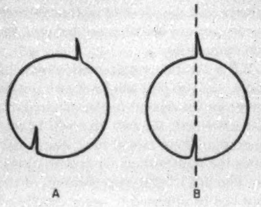


Figure 1. Circuit for producing circular or elliptical trace on a cathode-ray tube

A—Phase-splitting circuit
B—Vector diagram

Paper 53-128, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted April 28, 1952; made available for printing November 28, 1952.

Dr. M. Abdel-Halim Ahmed is with Fouad I University, Giza, Cairo, Egypt.



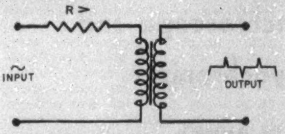


Figure 2 (left). Circular trace appearing on screen of cathode-ray synchroscope

Figure 3 (above). Circuit for pulse generation

teresis and eddy-current losses and the curvature of the magnetization curve of the core, there will be a time-lag displacement in the positions of the pulses as well as a distortion in their shape, and the secondary voltage wave form will be as shown in Fig. 4(*C*), which shows that the voltage rises rather slowly but decays sharply. The two peaks occur at 180 degrees, but lag a certain angle (about 30 degrees in the actual transformer used) behind the zero points of the original sine wave. Adjustment of the synchronization instant can evidently be made more accurately on the sharp side of the pulse.

It is evident that a magnetic-core material having an almost rectangular hysteresis loop with high permeability and low coercive force will give pulses approximating the desired ideal shape. To obtain pulses of good shape and sharp peaks, the number of the necessary ampere-turns should be more than about 10 times the ampere-turns necessary for producing the maximum flux density in the usual design of a normal power transformer using the same magnetic-core material.

When the two pips appear on the cathode-ray tube, confusion may result, especially when they are of large amplitudes, and it is preferable to suppress one of them. This can easily be done by any of the conventional methods, such as connecting a copper-oxide rectifier element across the secondary of the pulse transformer. The proper synchronization instant will then be when the remaining pip is on the Y axis in the upper part of the circle.

ADJUSTMENT OF CORRECT POSITIONS OF PIPS

If, at the correct instant for synchronization, the pips are required to be on the Y axis, a correction must be effected for the transformer time lag, in addition to the time advancement of the X-plate voltage ahead of the bus-bar voltage, which is inevitably produced by the phase-splitting circuit. A total correction angle of about 75 degrees is required. This is easily done by applying phase-shifting circuits before the phase-splitting circuit, and actually the phase-splitting circuit itself could be designed to incorporate the necessary correction. Many circuits have been successfully used, and the circuit shown in Fig. 5 was adopted in the final design of the apparatus. It has the following features:

1. Gives good wave forms of the voltages applied to the deflection plates. The resistance-capacitance circuit acts as a filter for the harmonics, if any.
2. It is easy to adjust the amount of phase-angle correction, and hence the correct position of the pips, without appreciably affecting the size of the trace, merely by varying resistance R.
3. It is easy to adjust the size and shape of the trace, without affecting the phase-angle correction appreciably, merely by varying R_2.

POWER SUPPLY FOR CATHODE-RAY TUBE

The power supply for the cathode-ray tube is obtained from the existing or bus-bar voltage, which is the one to be connected first to the apparatus; thus the circle would appear first. A conventional power supply circuit is used, employing a half-wave rectifier.

The high-tension secondary of the cathode-ray tube mains transformer may be used also for feeding the phase-splitting circuit. The phase-shifting circuit will also correct for any phase shift between the secondary and primary voltages of this transformer.

ROTATING-SPOT PRINCIPLE

The pulse signals, instead of being applied to the Y plates of the cathode-ray tube, may be applied to the grid of the tube, and the grid bias is adjusted so that the beam disappears when the pulse voltage falls a few volts below the positive peak value. In this case, instead of having two rotating pips, we get one rotating illuminated point, and the original circle will not be visible. The proper instant for synchronization is when the spot lies on the upper part of the Y axis.

FREQUENCY OF GENERATING STATION

If the synchroscope is adjusted for, say, 50-cycle operation, it will be found that it works well with 40 and 60 cycles. The only discrepancy that may appear is that the circle becomes an ellipse. This does not affect the operation of the apparatus in the least.

The characteristics of the pulse transformer may vary with frequency, giving different time-lag angles. This was found to be almost exactly compensated with the variation of the phase shift produced by the phase-shifting circuit. However, if greater accuracy is desired, one may vary the value of resistance R, Fig. 5.

THE EXPERIMENTAL SYNCHROSCOPE

The connection diagram for the developed experimental synchroscope is shown in Fig. 6. Switch S_1 is a 6-pole double-throw switch used to change the operation of the synchroscope from rotating pips (position 1) to rotating spot (position 2). Switch S_2 is a single-pole 3-way switch provided to enable the use of the synchroscope on either 40, 50, or 60 cycles. Both switches are added only for demonstration purposes and may be dispensed with in commercial apparatus.

Terminals 1 and 2 are for connection to the bus-bar voltage (110 volts), while terminals 1′ and 2′ are for connection to the incoming line. Both pairs should be connected to identical phases. Terminals 2*a* and 2′*a* are provided for use with 220 volts.

The cathode-ray tube is shielded against external electrostatic and magnetic effects. The experimental synchroscope is shown in Fig. 7.

CALIBRATION OF SYNCHROSCOPE

The calibration of the apparatus is done most easily by connecting terminals 1 and 1′ together, and terminals 2 and 2′ together, to one supply. Stationary pips or spot will thus appear, giving the exact positions for correct synchronization.

This procedure is also used when it is required to adjust the circuit elements to bring these positions of correct synchronization exactly on the Y axis.

FEATURES OF THE NEW SYNCHROSCOPE

1. Easy to make.
2. Easy to adjust and operate.
3. Cheap and durable.
4. More convenient; can be installed on the control desk.
5. Accuracy is the same as in normal synchroscopes, or better.
6. Not liable to errors, sticking, damage, etc.

The Cathode-Ray Automatic Synchronizer

The cathode-ray synchroscope is used when manual synchronizing of generators is desired. However where automatic synchronization is preferable, the cathode-ray automatic synchronizer will be found a simple and reliable gear.

The main parts of a cathode-ray automatic synchronizer are:

1. A cathode-ray synchroscope identical to that described previously and working on the rotating spot principle discussed before. The normal cathode-ray tube of the synchroscope is dispensed with, as it is replaced in the automatic synchronizer by a special tube. However, the normal cathode-ray tube may be used with the special tube, to serve as a visual monitor on the automatic synchronizer, or for manual synchronization when necessary.

2. A special cathode-ray tube, in which the beam current, as it traces the ellipse, is commutated by two electrodes replacing the screen of a normal tube. The idea of collecting a cathode-ray beam by such electrodes, instead of by the accelerating anode of the tube after having traced the desired phenomenon on the screen, was first suggested in 1938 by Colebrook,[1] who discussed a new principle for the generation of ultrashort waves, and since that time many devices have been suggested or utilized incorporating the same feature, but all were concerned with applications in the ultrashort wave field.

The application of this principle in the field of electronic measuring instruments and apparatus was proposed by the author in 1947;[2] several new types of electronic measuring instruments were presented, some of which used cathode-ray commutation. The application of the cathode-ray commutation principle in this field did not remain a mere proposal, but has been successfully verified experimentally during the past few years in our university.

3. Normal electromagnetic relays, which are actuated by the commutating-electrode currents of the cathode-ray tube, and which in turn cause the operation of the synchronization circuit breaker.

General Principle

A cathode-ray tube of the commutating-electrode type is used. The tube has two commutating electrodes A and B, Fig. 8, replacing the fluorescent screen. Through the use of circuits similar to those associated with the cathode-ray synchroscope, deflection voltages derived from the bus bars are applied to the X- and Y-deflection plates, while the pulses derived from the incoming line are applied to the grid of the tube.

In the absence of the pulses, the beam will trace a circle (or an ellipse) every cycle, and will be intercepted by electrode A, which subtends a very small angle θ, during a very small part of the cycle. The time of interception is equal to $(\theta/2\pi)T$, T being the time period of a cycle.

For the operation of the tube as a synchronizer, the tube is biased beyond cutoff, the beam current in this case will be zero, and we can only imagine a beam with zero current tracing a circle.

When the pulses are applied, the negative pulse will have no effect, but the positive pulse (occurring once every cycle) will cause a current to flow, provided the pulse has a sufficient peak value. This current will be collected by either electrode A or B, depending on the instant in which the positive pulse occurs; and the duration of the current will depend on the pulse width.

Assuming that the bus bar and incoming voltages are in phase and of equal frequencies (which are the required conditions for correct synchronizing), the positive pulse will occur whenever the beam traces electrode A. Hence electrode A will receive all the pulse currents, while the current in electrode B is always zero. Use is made of the current in electrode A to operate a relay which will cause the synchronizing circuit breaker to close.

If the two voltages are of equal frequencies but of different phase angles and the phase difference is greater than $\pm(\theta/2)$, all the pulse currents will fall on electrode B (assuming a very narrow pulse width), while the current in electrode A will be zero; therefore the relay incorporated with its circuit will not operate.

However, the relay will operate if the phase difference is less than $(\theta/2)$, and this is always allowable if the angle θ is small, as the incoming machine will be

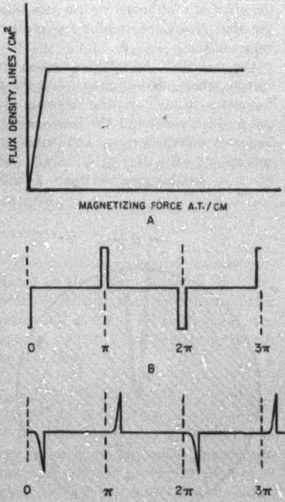


Figure 4 (left). Behavior of pulse generating transformer

A—Hypothetical magnetization curve
B—Ideal shape of pulse
C—Actual pulse wave form

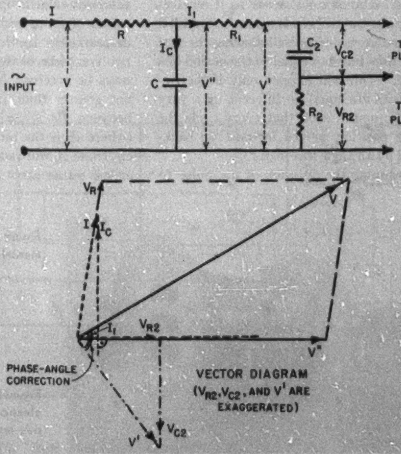


Figure 5 (right). Combined phase-shifting and splitting circuit for adjustment of pulse positions on the screen

$R = 50{,}000$ ohms
$R_1 = 500{,}000$ ohms
$R_2 = 12{,}000$ ohms
$C = 0.05$ microfarad
$C_2 = 0.25$ microfarad

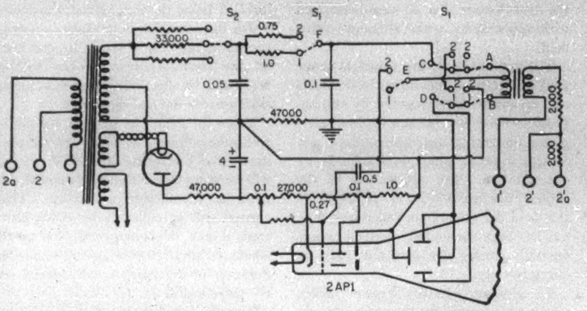


Figure 6. Connection diagram of experimental synchroscope (for demonstrations)

able to step into synchronism with the bus bars. It is therefore necessary that the angle θ should not be larger than what the synchronization conditions permit.

Now, if the two voltages are of different frequencies, the pulse will not occur when the beam is tracing a definite point of the circle, but the interception point will be moving forward or backward according to whether the frequency of the incoming voltage is smaller or larger. It will therefore be evident that only a few pulses fall on electrode A, while the majority will fall on electrode B. Bearing in mind that only one pulse occurs per cycle, the general shape of the currents in the electrodes will be as shown in Fig. 9.

Since it is required that the relay incorporated with the circuit of electrode A must not operate in this case, this relay must not be a quick-operating one, that is to say, it must not operate as it receives one pulse but will remain unoperated till it receives a second pulse following the first by a time interval equal to the period of a cycle. This will happen only if the frequencies are equal or different by a very small amount, such that every cycle the pulse position moves forward or backward by an angle less than θ.

Assuming the duration of the pulse to be negligibly small, it is evident that two (or more) consecutive pulses will be received by electrode A if the frequency difference (δ) is less than $[(\theta/2)/2\pi]f$, where f is the bus-bar frequency. If δ lies between $[(\theta/2)/2\pi]f$ and $(\theta/2)\pi f$, electrode A might receive one pulse or two consecutive pulses, according to the position of the first pulse on electrode A. For larger frequency differences only one pulse is received.

And it is always possible to effect synchronization when a small frequency difference exists (which case corresponds to the very slow motion of the pointer of an ordinary synchroscope), we can see that, also in this case, the angle θ has to be limited to correspond to the allowable frequency difference.

It is undesirable here that the pulse have a very narrow width, as the average current flowing in the relay must be of sufficient value to operate the relay. Also pulses of finite widths are more feasible in practice. In this case, the effect of two complete consecutive pulses will always be received by electrode A if δ is not greater than $[(\theta/2)/2\pi]f$. If δ lies between $[(\theta/2)/2\pi]f$ and $[(\theta-W)/2\pi]f$ (where W is the pulse width in radians), electrode A will receive the effect of two or less pulses according to the position of the first pulse on electrode A. For higher frequency differences, electrode A will never receive the effect of two pulses.

It is obvious that in all cases where electrode A receives two, or more, pulses, the two voltages, besides being of very nearly equal frequencies, are almost in phase at the time when the pulses are received, and hence the two necessary conditions for correct synchronization are both present.

The number of consecutive pulses received by electrode B (or missed by electrode A) will be $(f/\delta)\pm 1$ less the number of consecutive pulses received by electrode A.

ESTIMATION OF ELECTRODE ANGLE θ

It has been mentioned that for values of the frequency difference δ between $[(\theta/2)/2\pi]f$ and $[(\theta-W)/2\pi]f$, the average current of electrode A will lie between the values corresponding to one or two pulses. This does not mean that the relay will not operate, but actually the relay will remain unoperated as long as it receives a current less than that of two consecutive pulses. As the phase angle between the incoming and bus-bar voltages is continuously varying, this state of affairs will continue until the incoming voltage phase angle corresponds to the favorable condition giving two consecutive pulses to electrode A. Therefore no positive operation of the relay is to be expected in this region, especially if we bear in mind that the time which elapses until the relay operates is not strictly related to the value of δ (although we can generally say that it will be smaller for smaller values of δ).

The fact that the relay is certain to operate when δ is less than $[(\theta/2)/2\pi]f$ suggests a suitable relation between W and θ which eliminates the inconvenient frequency difference range, and that is by making $[(\theta-W)/2\pi]f=[(\theta/2)/2\pi]f$, or $W=\theta/2$. In this case electrode A will

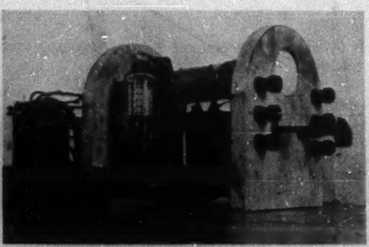


Figure 7 (left). The experimental cathode-ray synchroscope

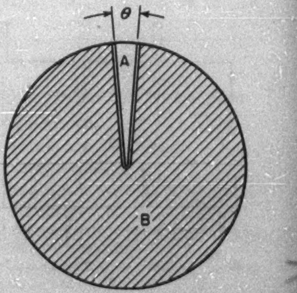


Figure 8 (right). Commutating electrodes of the special cathode-ray tube for the automatic synchronizer

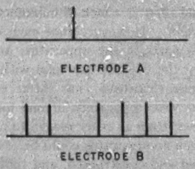


Figure 9. Commutating electrode currents

always receive two (or more) consecutive pulses when δ is less than $[(\theta/2)/2\pi]f$, and receives less than two pulses for larger values of δ.

The value of the angle θ will thus be given by: $\theta=(4\pi\delta)/f$ radians, where δ is the maximum allowable frequency difference for synchronization.

PULSE WIDTH

In the foregoing we were confined to ideal pulses of rectangular shape. However, the operation of the relay is concerned mainly with the shape of the current of electrode A rather than with the shape of the pulse voltage applied to the grid of the tube. The shape of this current, besides being of course affected by the shape of the pulse voltage, is also affected by several other factors, such as the bias voltage, the beam-current/grid-voltage characteristic of the tube, the focusing of the beam, its diameter and electron distribution, the diameter of the circle traced by the beam, the edge effects when the beam traverses one electrode to the other, secondary emission effects, etc.

Perhaps the easiest method taking all such effects into consideration is to trace the actual shape of the electrode current with normal operating values of the different parameters. A rectangular pulse can then be drawn having the same peak value and the same area as the actual pulse, and the width of this rectangular pulse can be taken as the equivalent width of the actual pulse.

SENSITIVITY

If I_p is the peak electrode current produced by the pulse, and W is the pulse width in radians, then if two consecutive pulses are received by electrode A, the average current flowing in the relay will be given by

$$I_{av}=2\frac{W}{2\pi}I_p$$

The beam current of normal cathode-ray tubes is usually small, but electromagnetic relays of high sensitivity can be easily designed, especially if good-quality core materials are used. Also special cathode-ray guns may be used to give high beam currents.

THE RELAY

The relay should be of the slow-acting type, operating with a current equal to the average electrode current I_{av} just mentioned, and having an operating time lag slightly less than two periods of the cycle. It is connected to electrode A through a capacitance-resistance circuit of suitable time constant to give the average current of the two received pulses, and zero current during the interval of the missed pulses. An interval will certainly elapse between the reception of the pulses by the electrode and the closing of the circuit breaker, and it may be found necessary to readjust the phase-shifting circuit, described in "Adjustment of Correct Position of Pips," to compensate for this new factor.

Since the average current flowing through the relay winding is affected by many factors and is not expected to be constant, some difficulty may be encountered in the design and adjustment of the delay time of the relay. A better procedure is to use a quick-acting, slow-releasing relay, which will operate as the first pulse current flows through it, but will have a release time lag larger than the period of one cycle. If the relay receives two consecutive pulses, it will remain operated for a period longer than the period of two cycles. This relay, when operated, will close the circuit of a second relay which is energized from a constant voltage supply. This latter relay shall be a slow-acting one, with an operating time lag equal to about the period of two cycles, and thus will only operate if the first relay receives two successive pulses. The second relay will operate the synchronizing circuit breaker.

FOCUSING THE BEAM

For proper design and operation of the apparatus, the cathode-ray beam should be well focused by using predetermined voltages applied to the tube focusing and accelerating electrodes.

If it is desired to focus the beam visually, this can be effected by viewing the collecting electrodes, since the incidence of the focused beam is made visible by fluorescence on the electrode surfaces.

Another suggestion is to cut out a sector of electrode B, and cover the glass of the tube behind that sector with fluorescent material, as in ordinary cathode-ray tubes. The collecting electrodes and the screen fluorescent material are made as far as possible in the same plane. The beam can easily be focused as it traverses this sector.

BASIC CIRCUIT

The basic circuit for the proposed cathode-ray automatic synchronizer is shown in Fig. 10. It is identical to that of the cathode-ray synchroscope, Fig. 6, used with the rotating spot principle, with the commutating-electrode circuits added.

Resistance R_k added in the cathode circuit of the tube serves to produce some cathode feedback as in ordinary thermionic valves. This will give a more linear grid-voltage/beam-current characteristic and, moreover, will increase the stability of the circuit.

It is also necessary that the collecting electrodes should have a potential higher than that of the accelerating anode, otherwise the electrons impinging on one of the electrodes will cause secondary emission, and electrons will travel back to be collected by the accelerating anode or by the other collecting electrode, especially if we take into consideration the

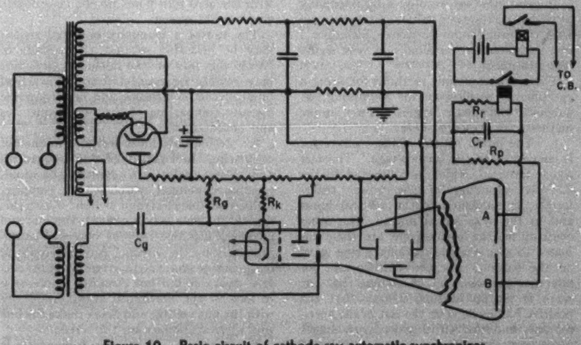


Figure 10. Basic circuit of cathode-ray automatic synchronizer

voltage drop in the circuit connected to the collecting electrodes.

Other methods to minimize secondary emission have also to be taken into consideration, among which we may mention the method adopted by Colebrook,[1] which uses electrodes in the form of buckets to prevent the escape of the secondary emission electrons from the interior electrode space. Again, the collecting-electrode material should be carefully chosen to cause as little secondary emission as possible.

Resistance R_p serves to make the circuits of the two electrodes nearly identical. Capacitor C_r acts as a by-pass to make the average current flow in the relay. The time constant of C_r and the relay circuit should be designed to the proper value. The resistance R_r may be dispensed with if the resistance of the relay is of sufficient value. It also helps to damp out any oscillations that may arise.

Conclusion

The developed synchroscope was tried out for the paralleling of the existing machines. It was found to work quite satisfactorily and to be quite convenient and reliable.

As regards the proposed automatic synchronizer, it has not been investigated experimentally, mainly because the cathode-ray tube is one of special construction and there is a lack of manufacturing equipment. However, since each of the various parts of the apparatus, and especially the cathode-ray tube with commutating electrodes, has been experimented with separately in one form or the other, there is every reason to believe that such an apparatus could be properly designed and constructed, and that good results could be achieved in practice with it.

References

1. ULTRA-SHORT AND DECIMETRE WAVE VALVES, **F. M. Colebrook.** *Wireless Engineer*, London, England, vol. 15, April 1938, pp. 198–201.

2. NEW TYPES OF ELECTRONIC MEASURING INSTRUMENTS, **M. Abdel-Halim Ahmed.** *Thesis*, Fouad I University, Giza, Cairo, Egypt, 1947.

◆

Discussion

William J. Kassimir (Atlantic Refining Company, Philadelphia, Pa.): The purpose of this discussion is to present other methods and procedures for accomplishing the objectives of Dr. Ahmed's paper, in a simpler manner and with apparatus that is commercially available and in no way special.

CATHODE-RAY SYNCHROSCOPE

The methods given here are those that I have been using for a number of years for setting the governors on small rotary d-c a-c converters and for setting the control rheostats on vibrator type d-c a-c converters. Both types of converters are energized from a d-c source and are used to supply 60 cycles to chart drive motor circuits on shipboard.

The converters were adjusted to 60 cycles by comparing their output with the 60 cycles supplied by the local power company using a cathode-ray oscilloscope for comparison. In this operation we were concerned only with the frequency and not phase-angle difference, but the methods discussed here take care of the phase angles as well.

In our operation we used a commercially available oscilloscope which was equipped with the usual input terminals including a test signal output terminal and sync signal input terminal.

There are numerous methods for using a cathode-ray oscilloscope for comparing frequencies and phase angles. Two simple methods are presented here:

1. The power supply of the oscilloscope is connected to the bus voltage. The test signal terminal, which is connected to the 6.3-volt secondary of the power transformer, is connected to the vertical input and to the sync signal input. The other controls on the oscilloscope are then adjusted to give a single stationary sine wave on the screen. For convenience and for reference purposes, let us assume that the wave is set up in such a way that the positive half-cycle is on the left of the vertical axis and the 180-degree zero-voltage-point goes through the zero-zero point of the co-ordinate graph on the screen.

The test signal is now disconnected from the vertical input terminal but still left connected to the sync signal input terminal. With this latter connection the horizontal sweep is synchronized with the frequency of the power supply.

The incoming generator output is now connected to the vertical input terminals of the oscilloscope and its wave form will appear on the screen. If its frequency is exactly the same as the bus frequency it will appear stationary on the screen by virtue of the fact that the horizontal sweep is synchronized with the bus frequency. If the frequency of the incoming machine is higher than the bus frequency, the sine wave will appear to move from right to left, and if the frequency of the incoming machine is lower than the bus frequency, the sine wave will appear to move from left to right.

The frequency and phase angle of the incoming machine will be correct, for closing the paralleling circuit breaker, when the sine wave on the screen is stationary and occupies the identical position of the sine wave that was initially put on the oscilloscope from the test signal terminal, that is, the positive half-cycle on the left and the 180-degree zero-voltage-point coincident with the zero-zero point on the co-ordinate graph on the screen.

The use of a manually operated switch may be desirable so that the voltage of either the bus or the incoming machine may readily be connected to the vertical input of the oscilloscope and hence provide a ready comparison of the two frequencies and phase angles.

2. Perhaps the most desirable method of comparing the equality of two frequencies and phase angles is to replace the aforementioned manual switch with a commercially available electronic switch. With this device, the sine waves of both the bus and incoming machine may be observed simultaneously on the screen. Any difference in frequency or phase angle is readily apparent. It is desirable but not absolutely necessary to lock in the horizontal sweep frequency with the bus voltage and hence make the bus sine wave stationary on the screen.

Exact equality of the two frequencies is obtained when the relative velocity between the two is zero. Both can be moving in the same direction at the same velocity when the two frequencies are equal. Since the two sine waves appear simultaneously, the phase-angle difference between the two is readily apparent.

By means of a control on the electronic switch, the two sine waves can be brought into exact overlapping alignment or may be displaced vertically from each other. The user may choose any degree of vertical displacement that he desires for comparing the equality of the two frequencies and phase angles.

This latter method, I believe, will appeal to most operators of generating stations because they are familiar with the appearance of an a-c sine wave and it will be easier for them to comprehend exactly what they are trying to accomplish. The use of a circular trace with a pip will not be understood easily by the average station operator unless he has had a course in electronics, without which he would be reduced to a mere automaton.

Another advantage of showing the two wave forms on the screen simultaneously is that the magnitude of the two voltages may be approximately compared. However, for this purpose the switchboard voltmeter will give a more accurate comparison of the two voltages.

AUTOMATIC SYNCHRONIZER

Dr. Ahmed describes a cathode-ray automatic synchronizer in which he uses a special cathode-ray tube with two so-called commutating electrodes. I leave out the word "cathode-ray" because I believe that the same results can be accomplished without a special cathode-ray tube, using only standard electronic parts.

Fig. 1 of the discussion shows a simplified functional diagram of my proposal. The circuit is intended to operate as follows: The plate of the triode receives power from the bus while the grid is connected to the incoming generator. With suitable grid bias the plate current through the relay can be held below the relay operating point until such time as plate and grid are both on their positive half-cycles, with sufficient current will flow through the relay to

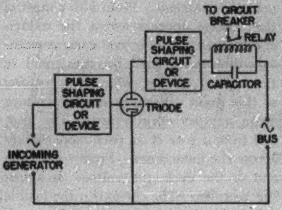

TO CIRCUIT BREAKER
RELAY
PULSE SHAPING CIRCUIT OR DEVICE
CAPACITOR
PULSE SHAPING CIRCUIT OR DEVICE
TRIODE
BUS
INCOMING GENERATOR

Fig. 1. Functional diagram of circuit used to provide relay operation when two separate a-c sources are in phase with each other

enable it to pull in. Hence, the necessary in-phase relationship is established, because when the two power sources are not in phase the relay will not pull in, owing to insufficient current.

So that the two power sources may be in phase within the allowable small-error angle, pulse-shaping circuits or devices are installed in each supply so that narrow pulses are obtained. If desired, saturable transformers, as described by Dr. Ahmed, may be used, or pulse shaping electronic circuits can be provided. If necessary, phase-shifting circuits can be incorporated.

Regarding the control relay, Dr. Ahmed's system of using fast- and slow-acting relays may be used, or a single relay may be employed, controlled from a pulse-dividing circuit so that several in-phase pulses may be received before the relay closes.

M. Abdel-Halim Ahmed: I agree with Mr. Kassimir that the methods he presented in the first part of his discussion are possible means for affecting synchronization. In fact, they represent two of the many basic methods of using a cathode-ray oscilloscope for the comparison of two sine waves, to which I may add a simpler method. This involves the application of one of the waves to the horizontal deflection plates of the oscilloscope tube, and the other to the vertical deflection plates. The trace will generally be an ellipse of varying shape. When the two waves are exactly of the same frequency and phase, the ellipse reduces to a straight line extending in the first and third quadrants.

The synchroscope presented in the first part of the paper is, however, another possible new means which has its own merits. I believe it is simpler than Mr. Kassimir's two methods, as it uses much simpler and far less expensive apparatus, all the components of which are commercially available. In the meantime, it gives a much simpler trace which is more convenient for observation and more comfortable for the operator, who is certainly familiar with viewing the rotating pointer of an ordinary synchroscope. The adjustment of the sharp rotating pips or spot is definitely more accurate, and the fact that the circular trace is much longer than the horizontal sweep adds to the better accuracy.

I do not agree with Mr. Kassimir that the operator should have a course in electronics before attempting to make use of the rotating pips or spot. On the contrary, this course is more necessary for an operator working with the whole bulk of equipment used in Mr. Kassimir's methods.

There is little to be said about the automatic synchronizer which Mr. Kassimir proposes in the second part of his discussion, as it actually works on the same principles involved in the automatic synchronizer presented in the second part of the paper, and entails all the main items in the latter. The difference lies mainly in the replacement of the commutating-electrode-type cathode-ray tube with an electronic tube, in view of the latter nowadays being commercially available. The paper presents a possible way for development, and in this respect I may mention that cathode-ray tubes similar to the commutating type that I have introduced are now commercially available in many forms for some other special purposes (for example, the ultrahigh frequency deflection tubes, the switch valves, the counter valves, etc). The new type of tube differs mainly in the collecting-electrode structure and therefore there will be no difficulty in developing such a tube. Moreover, this tube has many other applications in which it exhibits unique features. Such a wide scope of applications would justify the manufacture of such a tube.

A point that should be stressed is that the pulse generating device connected to the triode plate must deliver a relatively large instantaneous power during the pulse period, and its characteristics will not be identical with those of the one connected to the grid. It is therefore imperative that provision be made for correcting the discrepancy.

An Instability of Self-Saturating Magnetic Amplifiers Using Rectangular Loop Core Materials

S. B. BATDORF
NONMEMBER AIEE

W. N. JOHNSON
ASSOCIATE MEMBER AIEE

ACCORDING to the simple approximate magnetic amplifier theory, a plot of the average output current against the control current of a self-saturating magnetic amplifier has the shape of the left-hand portion of the hysteresis loop of the core material. Maximum amplification thus calls for the use of a core material having a rectangular hysteresis loop.

The use of rectangular loop core material increases the gain, but gives rise to an unanticipated operating instability sometimes referred to as "triggering." While the load current can be reduced smoothly from its maximum value by suitable adjustment of the control current, once having reached minimum value the load current cannot be increased again smoothly but jumps suddenly to a rather large value, as indicated in Figure 1.

This effect was noted by F. S. Malick in 1950 in an unpublished paper and attributed by him to any of three causes: an improper amount of positive feedback, an inductive load, or the use of rectangular hysteresis loop core material. The mechanism given for the instability when caused by the core material depended upon the difference between a minor and the major hysteresis loop; thus a given value of control current could result in either of two different output currents, depending on whether the material was operating on a minor or major hysteresis loop. The precise nature of the difference between minor and major hysteresis loops was not elaborated in detail, and appears in fact not to have been widely understood for rectangular loop material at the time the writing of this paper was undertaken. Because of this and be-

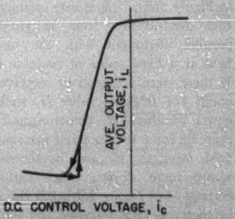
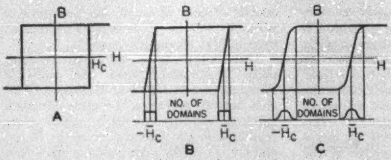


Figure 1. Transfer characteristic of self-saturating magnetic amplifier that triggers

---

Paper 53-18, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted September 5, 1952; made available for printing November 6, 1952.

S. B. BATDORF and W. N. JOHNSON are with the Westinghouse Research Laboratories, East Pittsburgh, Pa.



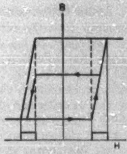


Figure 2 (left). B-H characteristics for three idealized models of a rectangular hysteresis loop material

Figure 3 (right). Magnetic behavior of idealized core material

cause a number of other phenomena such as leakage of rectifiers and eddy currents might be expected to influence triggering, the role of the core material in triggering remained in doubt.

The present investigation was undertaken to clarify the role of the core material. Representative cores were subjected to d-c measurements while being cycled in a manner similar to that occurring in magnetic amplifier operation. Later flux density–field intensity (B-H) traces were made continuously on the Cioffi recording fluxmeter.[1]

It was found that the behavior of rectangular loop core material is such that triggering will occur when it is used in a typical self-saturating magnetic amplifier circuit even if all other circuit elements behave ideally. Although triggering will, of course, be influenced by departures from ideal behavior on the part of other circuit elements, the present paper neglects these secondary effects and concentrates attention on the fundamental aspect of the phenomenon.

Theory

In rough approximation, a rectangular hysteresis loop material behaves as though it consisted of a large number of independent domains, each oriented either parallel or antiparallel to the magnetic field. If in such a material all domains had the same coercive force, the hysteresis loop would have vertical sides as shown in Figure 2(A). (It is here assumed that each domain flips from antiparallel to parallel alignment with the magnetic field at a fixed value of coercive force.) If there were a spread in coercive force about the average value H_c such that equal numbers of domains had their coercive force within equal increments of field strength H, the hysteresis loop would have sloping sides and sharp corners as in Figure 2(B). If the spread in coercive force followed the error frequency distribution, the hysteresis loop would have sloping sides and rounded corners as in Figure 2(C). The lower parts of Figures 2(B) and 2(C) show the number of domains as a function of the coercive force H_c. In the following analysis, the simple model shown in Figure 2(B) will be used, not with the thought that it represents accurately the physical phenomena concerned, but simply because it affords a convenient basis for a phenomenological formulation of the B-H relations of the material.

If a material of the type shown in Figure 2(B) is brought from negative saturation to some value of flux short of positive saturation and then returned to negative saturation, the hysteresis loop will be as shown in Figure 3. While the field is increasing, the material will follow the external hysteresis loop. Upon reversal of the field, the magnetic induction remains constant until the negative field attains a value corresponding to the lowest coercive force. Further increase of the field in the negative direction reverses the magnetization of the domains; this process continues until negative saturation is reached.

Consider next a core of such material in a half-wave self-saturating magnetic amplifier circuit of the type shown in Figure 4. The d-c control current provides a constant bias field (the choke is assumed to eliminate all feed back from the a-c circuit). If the bias current is in the negative saturation region, as at point a of Figure 5, the applied voltage will appear entirely across the load until point b is reached. From b to c most of the voltage appears across the inductance, till at point c the current-resistance drop in the load becomes equal to the applied voltage, which is decreasing at this point. From c to d the voltage again appears entirely across the load, and from d to e it is once more mainly across the inductance. From e until the end of the cycle the voltage is across the rectifier. The current in the load is shown in the lower half of Figure 5. The same sort of behavior, with negative saturation occurring during part of the cycle, will characterize any operating point a to the left of point e of Figure 5.

Consider next an operating point a', Figure 6, to the right of point e, starting again with a negatively saturated core. The cycle proceeds in the same manner as before except that at the end the material does not return to negative saturation but experiences a net gain in induction $a'f$, see Figure 6(A). During this and each succeeding cycle, the change in induction during the positive half of the cycle is given by the equation

$$(\Delta\phi)^+ = \frac{1}{N}\int_0^{\frac{T}{2}} (e_L - i_L R_L)\,dt$$

unless the occurrence of positive saturation prevents an increase of this magnitude. The decrease in induction during the negative half-cycle would be given by the same expression were it not limited by the fact that the presence of the rectifier prevents the field from dropping below that provided by the bias current. For the material under consideration both $(\Delta\phi)^+$ and $(\Delta\phi)^-$ remain substantially unchanged in successive cycles, so that in each cycle there is a net increase in induction

$$\Delta\phi = (\Delta\phi)^+ - (\Delta\phi)^-$$

Thus, after n cycles the core material will reach positive saturation where

$$n \cong \frac{\phi_s - \phi_e}{\Delta\phi}$$

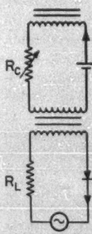


Figure 4 (left). Half-wave magnetic amplifier circuit

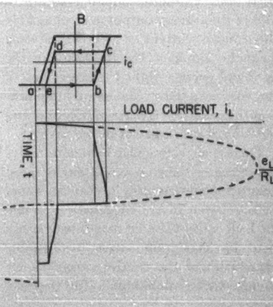


Figure 5. (right). Load current in half-wave amplifier when control current produces negative saturation

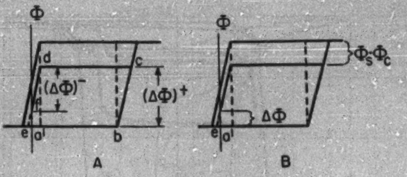


Figure 6. First cycle of ratcheting

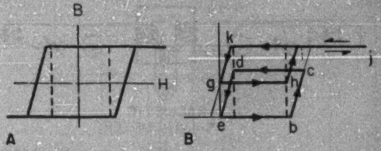


Figure 8. Magnetic characteristic resulting in ratcheting followed by triggering

see Figure 6(B). This situation may be summarized briefly by stating that as the bias current passes to the right of point *e*, Figure 6(A), the operating loop of the core material ratchets up to positive saturation. The final operating loop and the associated load current for the bias indicated in Figure 6 are shown in Figure 7. The sudden increase in load current from the value required to saturate the core to the value e_L/R_L corresponds to the firing of the amplifier. Firing does not occur when the bias is at point *e* even after ratcheting up to saturation because in the final loop saturation only occurs near the end of the cycle when e_L/R_L is equal to the saturation current for the core. Consequently, no discontinuity in the transfer characteristic is associated with the ratcheting, which normally occurs at the bias point *e*; that is to say, the ideal core material under consideration will not trigger in an ideal circuit after the fashion illustrated in Figure 1.

Now experimental evidence, to be discussed more fully later, indicates that a larger field is required to start reversal of magnetization when the core is initially saturated than when it is not. This suggests that the independent domain model employed up to this point should be amended to allow a small interaction resulting in nucleation phenomena. When the material is saturated there are no oppositely oriented domains present to nucleate reversal of magnetization, so that a larger field is required for this purpose than would be required for unsaturated material. Under these circumstances, the upper left-hand corners of all minor loops will be in a vertical line, shown dotted in Figure 8(A), while the upper left-hand corner of the major loop will be displaced to the left of this line. The bias required to start upward ratcheting in a material of this type will be such that when the ratcheting is complete, positive saturation will be reached early enough in the positive half-cycle of the supply voltage so that an appreciable voltage will appear across the load. This will be evident from a study of Figure 8(B). This figure shows the first cycle (*ebcde*) of the ratcheting operation as well as the final operating loop (*ghijkg*). In Figure 9, the hysteresis loop and the transfer characteristic are plotted with a common abscissa, to show the relationship between position of operating loop and output.

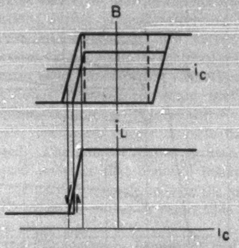


Figure 9. Relation between magnetic behavior and transfer characteristic

Before discussing the experimental part of this investigation, another result of the theory outlined should be mentioned which does not play an essential role in triggering, but can be seen in the data and serves to corroborate some aspects of the theory just outlined. The effect shows up as a peculiarity in the intermediate loops encountered in the upward ratcheting process.

Consider the changes in orientation of the domains during the first 2 cycles of ratcheting. In the first cycle, the material remains in a state of negative saturation in passing from *a* to *b*, Figure 10(A). At point *c*, half the domains are pointing one way, and half the other, those positively oriented being those of lowest coercive force as indicated in the figure. At the end of the cycle, point *f*, the domains are divided into three groups, those of low and high coercive force being negatively oriented and those of intermediate coercive force being positively oriented.

In the second cycle, Figure 10(B), the low coercive force negatively oriented domains are reversed in going from *g* to *h*. The next set of domains are already positively oriented, so no change in induction occurs as the field increases until point *i* is reached on the external hysteresis loop. The external loop is followed until the proper value of $(\Delta\phi)^+$ is attained at *j*, and the field is again reversed. It turns out that the maximum value of induction occurring in a given cycle determines the location of the horizontal portion of the right leg of the hysteresis loop associated with the succeeding cycle. This horizontal portion stretches from the minor to the major loop as shown in the figure.

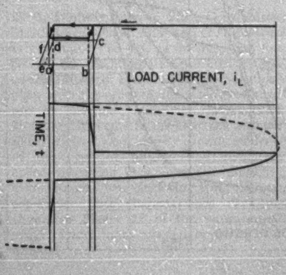


Figure 7 (left). Load current in half-wave amplifier when control bias is at point a'

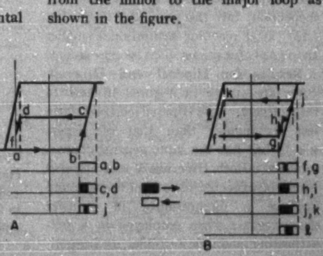


Figure 10 (right). Relation between hysteresis loops and domain behavior in accordance with theoretical model employed

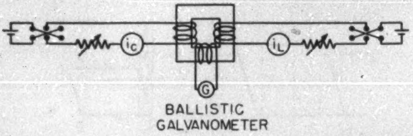


Figure 11 (above). Circuit used to measure magnetic characteristics of core

Experimental Procedure and Results

The characteristics of several toroidal cores were investigated using the circuit of Figure 11. First, the major hysteresis loop was determined using conventional d-c test procedures involving the ballistic galvanometer.[2] Then after choosing an appropriate bias current i_c, a change of flux $(\Delta\phi)^+$ was obtained by introducing a current i_L into the load circuit in several increments. After thus simulating the rise to maximum current in the self-saturating magnetic amplifier, i_L was reduced to zero in increments in order to complete the cycle. An attempt was made to keep $(\Delta\phi)^+$ constant in successive cycles, but this was done only approximately because it involved guessing the appropriate value of i_L each time.

The ratcheting of Hipernik V is shown in Figure 12. The dashed curve gives the major hysteresis loop. The B-H history during the ratcheting is shown by the solid curves for the case of a bias field of 0.138 oersted. The experimental results exhibit the main features discussed in connection with the theory: the upper left-hand corners of all minor loops lie approximately along one vertical line and the lower right-hand corners of minor loops lie along another; $(\Delta\phi)^-$ is substantially constant in successive cycles except when saturation is reached; and the right-hand sides of minor hysteresis loops have a nearly horizontal portion at the highest value of induction encountered in the previous cycle. In accordance with the theory, there is a slight horizontal deviation between the upper left-hand corner of the major loop and the vertical line joining the upper left-hand corners of the minor loops. This can only be detected upon rather close inspection, but the associated fact that $(\Delta\phi)^-$ after reaching saturation is smaller than $(\Delta\phi)^-$ for minor loops is very evident.

Results for Hipersil and Hipernik cores are shown in Figures 13 and 14 respectively. The Hipersil loop is slightly less rectangular than that of Hipernik V, but has the same general characteristics. Both Hipernik V and Hipersil would be expected to ratchet upward and produce a discontinuity in transfer characteristic when employed in a self-saturating magnetic amplifier, and experi-

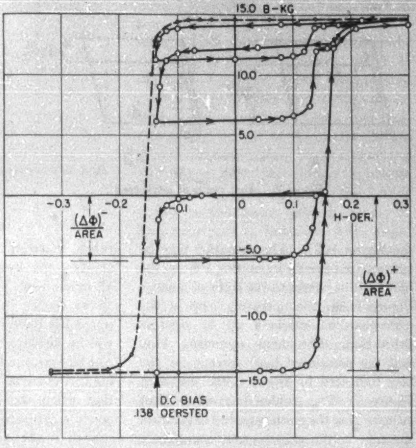


Figure 12 (right). D-c triggering of Hipernik V magnetic amplifier core 3/4 inch by 1 inch by 1/4 inch

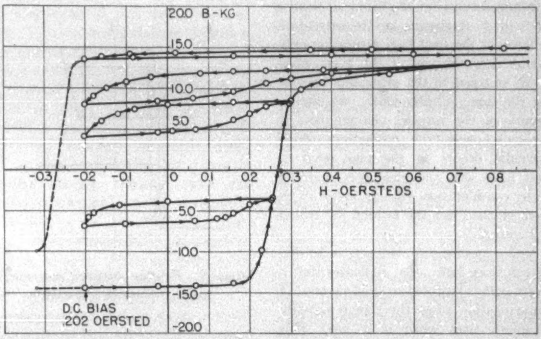


Figure 13 (above). D-c triggering of a Hipersil magnetic amplifier core 3/4 inch by $1\frac{1}{8}$ inches by 1/4 inch

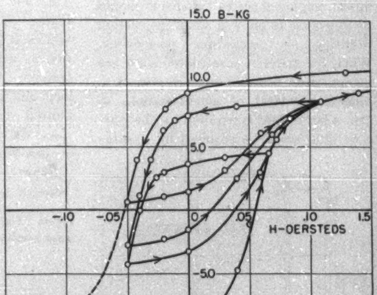


Figure 14 (right). D-c triggering procedure as applied to 0.002 inch Hipernik magnetic amplifier core 3/4 inch by 1 inch by 1/4 inch

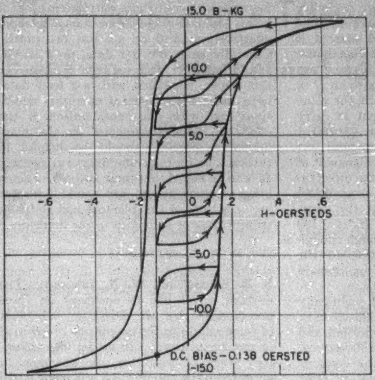
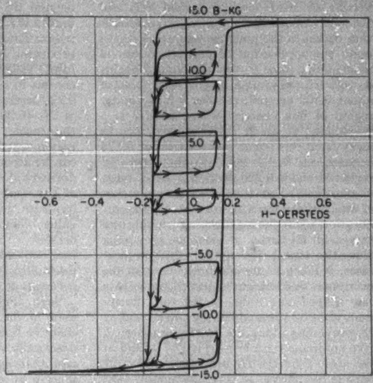


Figure 15 (left). D-c triggering procedure applied to a nonrectangular loop core material

Figure 16 (right). D-c triggering procedure applied to a Hipernik V magnetic amplifier 3/4 inch by 1 inch by 1/4 inch core in downward direction



ment confirms this result. The Hipernik loop is less rectangular than the other two and departs rather widely from the theory outlined. In particular, $(\Delta\phi)^-$ after reaching saturation is substantially the same as $(\Delta\phi)^-$ for interior loops. Accordingly, it would be expected to trigger weakly or not at all in a magnetic amplifier with ideal circuit elements. Experimentally it was found to trigger weakly.

Figure 15 shows similar experimental results for a nonrectangular loop core material. It is seen that $(\Delta\phi)^-$ keeps increasing as the operating loop rises, so that ratcheting to the top would not occur with this material, and there should be no discontinuity in the transfer characteristic. Experimentally, it was found that this core does not trigger.

Figure 16 shows a simulation of the cycling of a magnetic amplifier core during the return to negative saturation. It appears from this diagram that ratcheting downward should not occur. This may be regarded as a consequence of the horizontal portion on the left-hand side of each interior hysteresis loop, the existence of which can be explained in a manner similar to that for the horizontal portions during upward ratcheting. It can be shown that if such horizontal portions did not occur the material would ratchet down to negative saturation at the lowest bias which would prevent positive saturation. Experimentally a weak downward trigger is sometimes noted. The data of Figure 16 make it appear probable that nonideal behavior on the part of other elements in the magnetic amplifier circuit is responsible for such triggering.

The loop shown in Figures 15 and 16 were traced out using the recording fluxmeter developed by P. P. Cioffi.

Summary

On the basis of experimental data a simple phenomenological theory was developed to account for the B-H characteristics of rectangular hysteresis loop core material such as Hipernik V. It was shown theoretically that cores with such magnetic characteristics lead to triggering of self-saturating magnetic amplifiers. This conclusion was verified experimentally in the absence of secondary effects such as feedback, eddy currents, rectifier leakage, and so forth, by subjecting the core to magnetomotive force cycling simulating that encountered in the magnetic amplifier and measuring the resulting changes in magnetic induction. It was also verified by direct check that in a magnetic amplifier cores expected on the basis of theory to trigger actually do trigger and conversely.

References

1. A Recording Fluxmeter of High Accuracy and Sensitivity, P. P. Cioffi. *Review of Scientific Instruments* (New York, N. Y.), volume 21, 1950, pages 624–28.

2. Ferromagnetism (book), R. M. Bozorth. D. Van Nostrand Company, Inc., New York, N. Y.

◆

Discussion

D. C. Dieterly (Armco Steel Corporation, Middletown, Ohio): We who are interested in the magnetics art, or in magnetic materials, are grateful to the authors for discussing the way cyclic magnetization takes place with the types of minor loops of interest in magnetic amplifier applications. However, it seems to me that the authors' illustrations do not show clear-cut evidence of the type of phenomenon commonly called triggering in a magnetic amplifier, even though I am quite sure from my own experience that characteristics like that shown in Figure 12 of the paper would trigger under many normal operating conditions.

When one attempts to compare the d-c data shown in the authors' illustrations with a-c minor loops in a magnetic amplifier circuit, the comparison becomes difficult because of differences in ΔB amplitudes. Minor loops of such small ΔB amplitudes would hardly be employed in most magnetic amplifier applications. Commonly the voltage impressed across a half-wave element of a self-saturating magnetic amplifier will be such as to produce a ΔB amplitude practically between the positive and negative knees of the normal magnetization curve. If the authors had allowed their ΔH amplitudes to be slightly higher (as they would have been in normal a-c excitation in a typical magnetic amplifier circuit), instead of arbitrarily limiting them to a value sufficient to produce a ΔB amplitude only a fraction of that in a normal magnetic amplifier core, the cores would not have ratcheted through several cycles to arrive at the new stable minor loop condition for that bias value. The final minor loop in Figures 12 through 16 apparently was the stable loop for the bias shown, and under such restrictive ΔH and corresponding ΔB conditions the magnetization could hardly have progressed from the initial state to the final state except through several cycles of ratcheting, as shown.

It also occurs to me that the authors' illustrations might be misinterpreted by their readers, in light of the theory presented, because of failure to take into account the effect of shape parameters on the resultant loop. As an example consider Figure 12, which shows that as cyclic reversals are made on the normal loop and the flux changes from the negative saturation region to the positive saturation region upon the application of positive values of magnetizing force H, the induction B begins its upward swing at a magnetizing force of about 0.100 oersted, reaches a fairly linear region of flux change (where dB/dH is a maximum) at about 0.135 oersted, breaks away from this linear region at about 0.178 oersted, and levels off into the saturated region at about 0.220 oersted. Yet it must be remembered that the abscissas are values of magnetizing force H calculated from current flowing in the magnetizing winding and expressed in terms of oersteds prevailing along the mean path length l, which in most cases, is taken as the algebraic mean of the maximum and minimum path lengths in a ring shaped core. But in the core dimensions used by the authors the corresponding H_i prevailing along the inside periphery l_i (the minimum path length) is 1.17 H and at the same time the H_o prevailing along the outside periphery l_o (the maximum path length) is only 0.875 H. If we calculate the values of H_i and H_o corresponding to the H values previously mentioned as having been taken from Figure 12, we get the following:

	H_i	H	H_o
B starts upswing at	0.117	0.100	0.088
dB/dH curve quite high at	0.158	0.135	0.118
middle of high dB/dH range	0.182	0.156	0.137
end of high dB/dH range	0.208	0.178	0.156
B levels off at	0.257	0.220	0.192

These data indicate that the noticeable upswing in B occurs at $H=0.100$ oersted probably because the corresponding value of $H_i=0.117$ oersted is barely sufficient to cause significant numbers of domains along the inside periphery of the core to reverse their direction of magnetization. When the value of H reaches 0.135 oersted, H_i is 0.158 oersted (apparently a value where large numbers of domains along l_i reverse almost simultaneously) and H_o is 0.118 oersted (where significant numbers of domains along l_o will just begin to reverse magnetization). Then, when the value of H reaches 0.156 oersted, domains along the mean path l are reversing at the maximum rate, while only a few higher energy domains are being reversed by the corresponding value of $H_i=0.182$ oersted at the inside periphery and while significant numbers of low energy domains are being reversed along l_o where the value of H_o is 0.137 oersted. Further increase of H, along mean path length l, to 0.178 oersted has caused reversal of all except a few reluctant domains along that path; but along the inside periphery l_i where the magnetizing force is 0.208 oersted the domain reversals are negligible in number, yet at the outside periphery the magnetizing force H_o of 0.156 oersted is producing domain reversals at the maximum rate. When the magnetizing current increases to a value corresponding to an H of 0.220 oersted along the mean path, the value of $H_i=0.257$ oersted is sufficient to carry the inside periphery into the saturated region where further increase in total flux can be accomplished only by realignment of domains along a path closer to the direction of the applied field than was accomplished by mere domain reversals. At this same value of current the value of H_o is 0.192 oersted, a value which appears high enough that no appreciable contribution to the total flux is now being made by domain reversals at the outside periphery.

Such analysis indicates that practically none of the domains of this material have coercive forces less than 0.120 oersted and practically none higher than 0.190 oersted, and that substantially all of the domains reverse at coercive forces extremely close to 0.155 oersted. Actually the material of this core must have had extremely rectangular characteristics. The seemingly linear slope shown between 0.135 oersted and 0.178 oersted in Figure 12 appears to be due solely to the fact that l_o was 33 per cent longer than l_i, so that between these values of H (along path l) the maximum domain reversals begin at the inside periphery of the core and progress to the outside periphery as the true H along any path reaches about 0.155 oersted. If the outside-diameter–inside-diameter ratio of this core had been nearer 1.00, such a core of this material would have had a higher maximum dB/dH than in Figure 12; and in a core of larger outside-diameter–inside-diameter ratio than used here would have had a loop still more sheared than in Figure 12. Then too, if the outside-diameter–inside-diameter ratio of this core had been considerably closer to 1.00, most of the horizontal part at the upper right-hand corners of the intermediate minor loops would have disappeared.

It is true that there is some connection between extreme rectangularity of d-c hysteresis characteristics and the likelihood of triggering trouble in a magnetic-amplifier, but even the most rectangular material characteristics will not trigger if certain circuit characteristics are present which affect the a-c loop characteristics in such a way as to minimize triggering. On the other hand, I have noted cases of cores which trigger upward and downward and yet do not have highly rectangular d-c loops, and which by the authors' d-c techniques would not be suspected of triggering. Whether a core triggers or not depends primarily on its a-c loop characteristics under operating conditions.

There is much yet to be learned about the factors which contribute to triggering, but it is my opinion these factors may be more effectively studied by a-c test procedures. Carefully made d-c test data, such as Mr. Batdorf and Mr. Johnson have presented in this paper, combined with a-c loop data, both made under normal operating conditions and with due consideration of the effect of shape parameters on the magnetization process, should be helpful in getting a better understanding of triggering. It is to be hoped that these authors, or some other research group, will conduct further investigations of this phenomenon and contribute more of such data to the literature.

S. B. Batdorf and **W. N. Johnson:** The authors would like to comment on three points raised by Mr. Dieterly. The first of these relates to the magnitude of ΔB used in our experiments. In an initial exploratory study of triggering, the phenomenon appeared to be absent if ΔB was made equal or almost equal to $2B_s$, but was present for smaller values of ΔB no matter how small this might become. The ΔB used in the investigation was therefore chosen for convenience, without reference to whether it was representative of values normally used in amplifier operation. So far as the number of cycles of ratcheting is concerned, however, this is not very directly related to the size of ΔB. In fact, we have many times observed an upward ratchet take more than 5 seconds from start to finish at 60 cycles.

The second point raised by Mr. Dieterly is a suggestion that the steepness of the static hysteresis loop is intimately related to the ratio of external to internal diameter of the core. This we believe to be the case; we made no point of it in the paper, however, because we were not prepared to elucidate this relationship fully.

The third point is Mr. Dieterly's opinion that the factors which contribute to triggering may be more effectively studied by a-c than d-c test procedures. This is really a question of philosophy of research. Our belief is that to arrive at a fundamental under standing of what is going on in triggering as opposed to a quick "fix," it is desirable to separate the factors involved. The d-c test data, as we noted in the paper, permitted a study of the core behavior in a half-wave amplifier divorced from the complicating effects of eddy currents, leakage of rectifiers, inadequate voltage regulation, feedback to the control circuit, and so forth. Such a study will, of course, be an incomplete account of the behavior of the magnetic amplifier until it is supplemented with studies of the effects which have been neglected; however, the principle of separating effects for individual study we believe to be generally sound, and in many cases indispensable for real progress.

The Influence of a Transverse Magnetic Field on an Unconfined Glow Discharge

W. D. McBEE
ASSOCIATE MEMBER AIEE

W. G. DOW
FELLOW AIEE

A 3-KW unconfined glow discharge in air at pressures from 0.3 to 10 millimeters of mercury has been observed in the presence of transverse magnetic flux densities from 0 to 7,000 gausses. Effects observed as the magnetic flux density is increased include:

1. The voltages of the positive column and the anode fall space exhibit very pronounced increases, the former by a factor of at least 100 between zero magnetic field and 6,000 gausses; see Figure 6.
2. The cathode fall of potential exhibits an initial rapid decrease, then a levelling off; in combination with item 1 this results in a minimum value of total discharge potential at an intermediate magnetic flux density; see Figure 6.
3. A transverse wind, directed according to the motor rule, is generated by the positive column plasma; see Figure 4.
4. The equipotentials become skewed, passing diagonally through the discharge; see Figure 5.
5. Cathode fall space phenomena, including the negative glow and the Crookes and Faraday dark spaces, become very much compressed; see Figure 2.
6. The positive column and the anode regions become much more luminous because of their greatly increased gradients; see Figure 2.
7. The sputtering rate increases greatly.
8. Random potential fluctuations, see Figure 7, increase in amplitude, and their characteristic frequency first increases, then decreases.
9. Movement of cathode fall electrons parallel to the magnetic field produces a sharply delimited hollow shell of discharge resembling an axial extension of the cathode post; see Figure 2.

An approximate analysis of the underlying discharge phenomena shows that as the magnetic flux density is increased, the ion drift velocity is only moderately affected, whereas the electron drift velocity is greatly reduced in magnitude and becomes nearly perpendicular to the electric field. Qualitatively, this behavior can be related to many of the experimentally observed phenomena.

The work reported in this paper represents a portion of an initial exploration of the effect of a strong transverse magnetic field on a gaseous discharge at moderately low pressures.[1] The experimental objective was to observe and measure the effect of the magnetic field on as many discharge characteristics as possible. The theoretical aim was an approximate analysis for use as a guide in interpreting and correlating the experimental observations.

A glow discharge was selected for study, partly because its characteristics in the absence of a magnetic field are well known[2, 3] but also because a large difference between positive ion and electron trajectories can be obtained at typical glow discharge pressures by using convenient values of magnetic flux density.

Experimental Observations

The discharge domain is shown in Figure 1. This assembly is immersed in an atmosphere of dry air in a vacuum chamber. Note that the nonconducting boundaries of the discharge are unconfined horizontally. Thus the discharge is free to assume a self-imposed contour in a horizontal plane.

Observed attributes of the discharge were:

1. Visible form and color.
2. Cathode sputtering.
3. Wind (a result of the magnetic field; to be defined later).
4. Volt-ampere characteristics of an auxiliary probe electrode.
5. Time-average discharge potential differences and distributions.
6. Discharge potential variations with time.

These were studied as functions of magnetic flux density, pressure, and discharge current. The magnetic flux density ranged from 50 to 7,000 gausses, the pressure from 0.30 to 10 millimeters of mercury, and the current from 0.05 to 2.5 amperes. As complete data are available elsewhere,[1] only data showing the effect of magnetic field variations at a pressure of approximately 1.2 millimeters and a discharge current of about 0.6 ampere will be presented here.

Figure 2 shows the discharge. The side view is reflected from a mirror to the left of cathode. At negligible magnetic flux density, Figure 2(A), the luminous portion of the discharge consists entirely of the bluish-white negative glow, except for a few small anode spots. Increasing the magnetic flux density results in the following changes, illustrated in Figures 2(B), 2(C), and 2(D):

1. The negative glow is compressed horizontally into a thin circular tube of intense luminosity.
2. The anode spots become more numerous and coalesce to form a ring or rings.
3. The Faraday dark space decreases in extent and a pink positive column appears between cathode and anode. This region is usually striated for moderate fields (200 to 800 gausses); see Figure 2(B). However, as the flux density is increased further, it becomes continuous and brighter, and it assumes a contour in the horizontal plane similar to the shaded area of Figure 5.

Sputtering from the copper cathode is negligible in the absence of a magnetic field but increases rapidly with increasing magnetic flux density. This is evidenced by copper deposits on transite sheets adjacent to the cathode and by the appearance of green color due to excited copper vapor in the region around the cathode.

The presence and direction of the wind[4] caused by the discharge were observed by suspending Fiberglas streamers at various positions around the discharge. With increasing magnetic field these streamers became displaced from their vertical rest position, ultimately becoming horizontal and fluttering as though in a turbulent wind. The wind direction can be reversed by changing the polarity of either the magnetic field or the discharge, but remains the same if both are changed. The wind direction is generally mutually perpendicular to the magnetic field and the current; in more detail, it is always at right angles to the boundary of the discharge. Its sense is given by the motor rule. The approximate velocity of the wind was measured by observing the angular deflection of a small, rectangular mica vane mounted as shown in Figure 3. Wind velocity is plotted as a function of magnetic flux density in Figure 4; absolute values given there may be in error by an appreciable factor either way, but the relative behavior is believed to be correctly indicated.

An attempt was made to obtain information by means of an auxiliary probe

---

Paper 53-154, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted October 22, 1951; made available for printing December 23, 1952.

W. D. McBee is with the Sperry Gyroscope Company, Great Neck, N. Y., and W. G. Dow is with the University of Michigan, Ann Arbor, Mich.

The authors wish to acknowledge the very important contributions made by H. C. Early in relation to the experimental arrangements and criticisms of the manuscript.

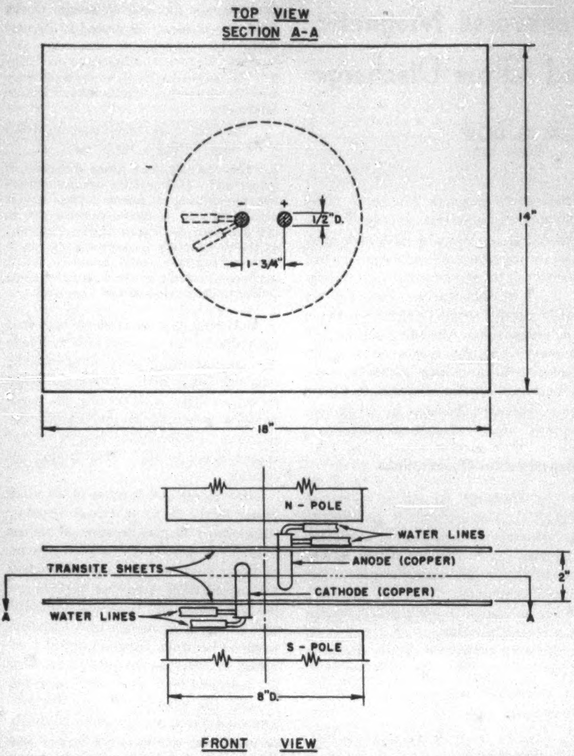


Figure 1. Electrodes, magnetic poles, and insulating discharge boundaries of the experimental apparatus

electrode.[2, 5–7] A disk, 0.020 inch in diameter, surrounded by an equipotential guard ring, was used as a probe. Probe volt-ampere characteristics were taken with the probe surface both parallel and perpendicular to the magnetic field. Because of large variations in discharge potential (discussed later), each point on the probe characteristic represents the average current collected over a range of potential differences between probe and discharge. Therefore, only wide plateaus of current saturation are significant. Plateaus indicative of ion current saturation were observed on all the characteristics. However, saturated electron current plateaus were observed only with the probe surface parallel to the magnetic field and with magnetic flux densities in excess of 1,000 gausses. The ratio of saturated electron current to saturated ion current decreased from 10 at 930 gausses to about 2 at 5,200 gausses. Usual values for this ratio in the absence of a magnetic field exceed 1,000.

The remaining data deal with discharge potentials. Large and rapid fluctuations in plasma potential prevented making plasma probe potential measurements in the usual way.[7] Therefore zero-current probe measurements were made to obtain the data for Figures 5 and 6. The ideal zero-current probe potential would differ from the true plasma potential by about the difference in the electron and ion mean energies (in electron volts), that is, by perhaps as much as 10 volts, which is not significant with the gradients of 200 or more volts per centimeter existing.

However, a probe has a rectifier type of characteristic[8] and a finite capacitance; therefore, depending on the magnitude of the capacitance and the plasma potential fluctuation frequency, a zero-current probe's potential may not follow the plasma potential faithfully during a rapid rise to a fluctuation crest, but will follow the plasma potential on the downswing. Therefore the d-c zero-current probe potential may be lower than the true average plasma potential, perhaps by several times 10 volts. This difference will, however, be consistent in trend throughout the measurements. If its presence is recognized, it need not invalidate interpretations of data based primarily on relative values of d-c zero-current probe voltages.

Thus all of the probe potentials used for Figures 5 and 6 are lower than true plasma potentials, but by consistent amounts, perhaps as much as 50 volts, but probably not more than 100 volts. However, relative values and equipotential shapes are believed to represent the true situation reasonably correctly. Probably the true anode fall of potential is substantially less and the true cathode fall of potential somewhat greater than indicated in Figure 6. This error will be a maximum at the largest B, and will decline to insignificance for small B.

A 0.005-inch-diameter tungsten wire floating probe having a 1/8-inch exposed length was used to obtain the horizontal-plane potential map of Figure 5. Note the skewing of the field, the large potential gradient in the positive column, and the divergent fanning-out of the equipotentials below the lower boundary. Equipotentials near the anode may have true potential values considerably higher than the indicated zero-current probe potentials because of the probe's rectifying characteristic. Thus the steep gradient near the anode may be partially an instrumentation effect.

The total discharge potential difference consists of the cathode fall, the positive column drop, and the anode fall, distinguished experimentally by placing a probe (number 1) 1/8 inch from the cathode and another (number 2) 1/10 inch from the anode. Figure 6 shows illustrative probe and cathode potentials relative to the anode, as functions of magnetic flux density. With increasing B, the positive column voltages increase, whereas the cathode fall decreases, causing a minimum in the total voltage curve.

Potential fluctuations were recorded by photographing the trace of a floating probe's potential on a single-sweep cathode-ray oscilloscope; see Figure 7. The relaxation character of the recorded voltage variations probably results from the probe following plasma potentials faithfully on the downswing, but not on the rise. In general, the fluctuation amplitude increases with both pressure and magnetic flux density. The observed frequency decreases with increased pressure. At low pressures, frequency increases with flux density, but a maximum frequency at an intermediate flux density was observed at the higher pressures. Details of changes in observed frequency may be influenced by changes in the probe volt-ampere characteristic, which affects the faithfulness of response to a potential rise.

Approximate Theory of Particle Mobilities

To indicate trends in interpreting the observed phenomena a roughly approximate mobility theory is useful.

Let

$\mathbf{E}=E_x\mathbf{i}=$ electric field intensity, volts per meter (1A)

$\mathbf{B}=B_y\mathbf{j}=$ magnetic flux density, webers per square meter (1B)

$\xi_0=u_0\mathbf{i}+v_0\mathbf{j}+w_0\mathbf{k}=$ initial velocity of a charged particle trajectory, meters per second (1C)

$L=$ mean free path (1D)

$t_L=$ Mean-free-path time (1E)

$\epsilon_0=\dfrac{mc_0^2}{2q}=$ initial energy of a particle, electron volts (1F)

$m=$ particle mass in kilograms; q = particle charge in coulombs (1G)

The x-z plane is the domain of the complex variable $r=x+jy$ (2A)

The x-z projection of the velocity, also complex $=c=u+jv$ (2B)

For mutually perpendicular electric and magnetic fields, the familiar cycloidal x-z projection of the trajectory, Figures 8, 9 possesses

$$\text{angular velocity:}\quad \omega=B_y\frac{q}{m} \tag{3A}$$

$$\text{generating radius:}\quad R_g=\frac{1}{\omega}\left|w_0-\frac{E_x}{B_y}-ju_0\right| \tag{3B}$$

$$\text{rolling radius:}\quad R_r=\frac{1}{\omega}\,\frac{E_x}{B_y} \tag{3C}$$

With $B>500$ gausses and E/B as in the experiments, suitable relative magnitudes are

For electrons with $\xi<10$ electron volts, $R_g>5R_r;\ \omega t_L>>1$ (4)

For ions with $\xi<3$ electron volts, $R_g\simeq R_r$ $\left(\text{that is, } |c_0|<<\dfrac{E}{B}\right);\ \omega t_L<<1$ (5)

From equation 4, an $\xi<10$ electron describes during one free path many cycles of a large-loop cycloid, Figure 8;[5] therefore, the primary motion is a drift at velocity E/B along equipotentials (perpendicular to both E and B). At each collision the generating radius and phase change, but not the E/B drift. On the average, each collision results in a shift of the cycloid's base line in the direction of the electric field force; this is the mechanism of the slight electron drift in the field force direction, when B is large.

Under conditions of equation 5, the typical ion has an initial velocity very modest relative to the E/B velocity. It therefore starts after each collision from a point at or near the cusp of the common cycloid (middle, Figure 9 curve), or at a comparable point on either of the other Figure 9 curves. Thus, on the average, its starting motion is primarily in the direction of the electric field force. Because each free path is a very small fraction of a cycloid's cyclic length, its whole travel during one free path will be chiefly in the direction of the starting motion. Thus ion drift takes place primarily in the direction of the electric field force, with only a modest E/B direction component.

Thus electrons move chiefly along equipotentials, whereas ions have their major drift in the electric field force direction. These contrasting properties dominate the gross nature of the discharge, as described by Figures 5 and 6, and result in the Figure 4 wind.

The general correspondence between the Figure 5 pattern and expected ion and electron behavior, discussed later, suggests that the Figure 7 potential fluctuations do not alter in a major way the general relative magnitudes of the average values of the components of ion and electron drift. They do, however, indicate large statistical departures from the averages, and may markedly affect the absolute values of the averages.

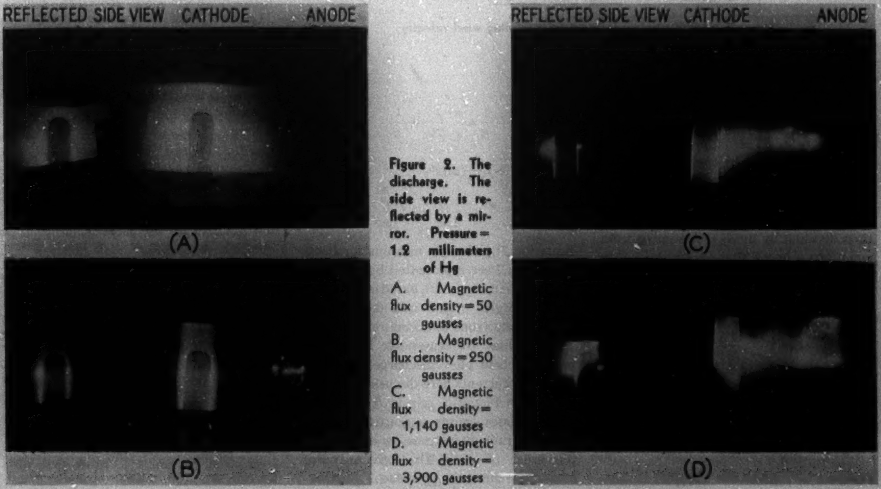


Figure 2. The discharge. The side view is reflected by a mirror. Pressure = 1.2 millimeters of Hg

A. Magnetic flux density = 50 gausses
B. Magnetic flux density = 250 gausses
C. Magnetic flux density = 1,140 gausses
D. Magnetic flux density = 3,900 gausses

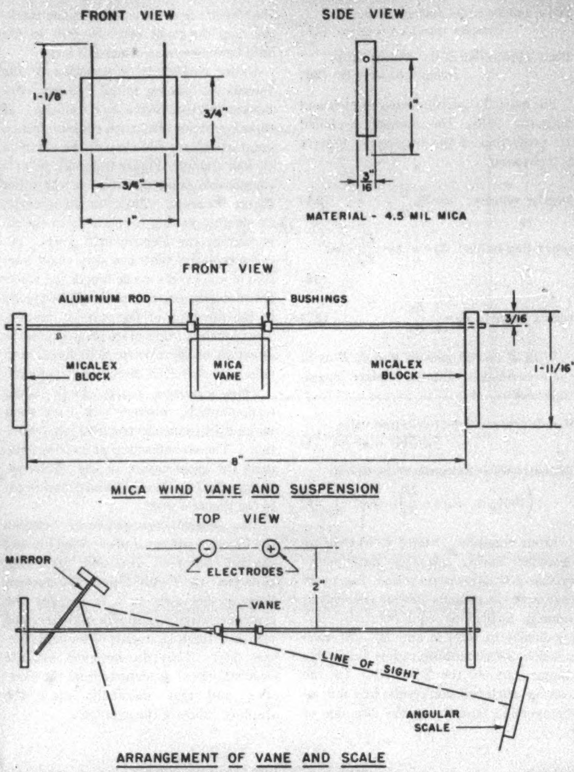


Figure 3. Mica vane arrangement for measuring wind velocity

Rather well-known "complex mobilities" can be derived from the velocities in the cycloidal trajectory, for a modest electric field. The approximate drift velocities are found by taking the line integrals of velocity, as a function of ωt from 0 to ωt_L, and as a function of the initial velocity direction from 0 to 2π, then dividing by $2\pi\omega t_L$. Assuming from equations 4 and 5 that $\omega t_L << 1$ for ions and $\omega t_L >> 1$ for electrons, drift velocities are found to be

$$\text{electrons: drift velocity} = \frac{1}{\omega t_L}\frac{B}{E} + j\frac{E}{B} \tag{6A}$$

$$\text{ions: drift velocity} = \frac{1}{2}\frac{E}{B}\omega t_L + j\frac{1}{6}\frac{E}{B}(\omega t_L)^2 \tag{6B}$$

Dividing by E gives the "complex mobilities"

$$g_e = \frac{1}{\omega t_L B}i + \frac{1}{B} = \frac{m}{q}\frac{1}{B^2 t_L} + j\frac{1}{B} \tag{7A}$$

$$g_i = \frac{1}{2}\frac{q}{m}t_L + j\frac{1}{6B}(\omega t_L)^2$$

$$= \frac{1}{2}\frac{q}{m}t_L + j\frac{q^2}{6m^2}Bt^2{}_L \tag{7B}$$

Equation 7(A) has no meaning for $B=0$. Equations 7 are represented by the solid-line curves in Figure 10, based on constant mean energies of 1.0 and 0.6 electron volts for electrons and ions respectively. These values are reasonable for a zero magnetic field glow discharge, for pressures and currents here discussed. However, from Figure 6, raising B from 0 to 6,000 gausses causes a positive column average gradient increase by a factor of the order of 100. This obviously increases the mean energy values. Furthermore, the mean free paths necessary for the Figure 10 evaluations are imperfectly known and will be affected by the air temperature rise that must accompany the great increase in energy input to the discharge between 0 and 6,000 gausses. Thus Figure 10 should be thought of merely as exhibiting certain relative-value general trends.

However, in support of the general utility of Figure 10, the equation 4 and equation 5 mean energy limitations to less than 10 and 3 electron volts respectively are probably valid at least to 6,000 gausses. For the electrons, rise to a greater ξ than about 10 causes excitation or ionization that drops ξ again to a low value. There are reasons of the self-consistency type to believe that the ion mean energy (ξ) in the steep magnetic plasma gradient of value E will vary directly as E, just as for ions in a strong nonmagnetic plasma.[5] On this basis, for expected free paths and Figures 5 and 6 gradients, ion mean energies should rise to the order of 1.5 electron volts at 6,000 gausses; this is still within the $\xi<3$ limit in equation 5, used for equation 7.

The equation 7 mobility components are functions of t_L; in general $t_L = L/c$ and c^2 varies in proportion to ξ; therefore t_L^2 varies roughly inversely as ξ. Use of this and the previous paragraph's concepts as to ξ permits comments as to expected deviations from Figure 10 at large values of B as follows:

1. *The imaginary component of g_i.* From Figure 6 the plasma gradient increases, nearly linearly, about one hundred-fold between $B=0$ and $B=6{,}000$ gausses. Therefore, with t_L^2 varying inversely as ξ, and ξ for ions varying as E, t_L^2 should for large B vary about inversely as E. Therefore, from equation 7(B) and Figure 6, the imaginary component of g_i should at 6,000 gausses be less than the Figure 10 values by perhaps a factor of 100.

2. *The real component of g_i.* From equation 7(B) the real component of g_i varies as t_L, therefore roughly inversely with $\sqrt{E}$. Therefore between 0 and 6,000 gausses the real component of g_i should decrease to a value lower by about a factor of 10 than at $B=0$. Thus for all magnetic fields used, the real component of g_i is much greater than the imaginary component of g_i. In this relative magnitude respect, Figure 10 leads to the correct concept, although magnitudes are far from correct for large B.

3. *The imaginary component of g_e.* Because the imaginary component of g_e does not involve the mean energy, it should vary roughly as indicated in Figure 10, except for a more or less abrupt drop to zero at $B=0$; equation 7(A) is not valid at $B=0$.

4. *The real component of g_e.* Because the electron mean energy may reasonably increase from perhaps 1 to 10 electron volts between $B=0$ and $B=6{,}000$ gausses, due to

the major general increase in energy input to the discharge, the real component of g_e must from equation 7(*A*) be greater at 6,000 gausses than Figure 10 indicates. The suggested increase in ξ by roughly ten times would indicate an increase over Figure 10 values by roughly ten times. However, indirect evidence,[7] related to random plasma voltage fluctuations, suggests that at 6,000 gausses the real component of g_e may be greater than the Figure 10 value by a factor between 10 and 50. This still keeps the real component of g_e substantially below the imaginary component; the shape of the equipotentials in Figure 5 is consistent with this point of view and would not be consistent with a real component of g_e larger than the imaginary component at 6,000 gausses.

These comments explain the dashed-line arrows in Figure 10, suggesting estimated values more appropriate at 6,000 gausses than those based on equation 7. They also indicate that for large *B* the major ion flow will be at right angles to equipotentials and the major flow of electrons along equipotentials.

Cathode and Anode Fall Spaces

The data interpretations given in this paper are consistent with the entire experimental study[1] only partially presented here.

Electron emission results from bombardment of the cathode by positive ions[2, 3, 5] accelerated through the cathode fall of potential. Because the cathode fall space is thin and has a very strong electric field, falling-in ions start their fall at a trajectory point corresponding to a cusp of the Figure 9 cycloid and reach the cathode within one or two free paths. Thus in the cathode fall space, the ion motion is not seriously affected by the magnetic field directly.

In contrast, with a strong magnetic field, the electrons experience a field configuration like that in a high-vacuum diode magnetron. The electrons can escape from rapid circulation around the cathode only by losing kinetic energy. Such loss of kinetic energy occurs only at exciting and ionizing collisions. The strong luminosity of the cathode fall region suggests that many such collisions must occur prior to escape from the tight circulation.

Because emitted electrons circle close in to the cathode, most of the cathode fall space ion generation occurs very near the cathode. This causes a marked contraction of the cathode fall space, together with an increase in luminosity as the magnetic field increases; see Figure 2. Because the ions are generated near the cathode, they suffer fewer collisions and therefore lose less energy in traversing the cathode fall space than in a nonmagnetic glow discharge. Therefore, the cathode fall of potential declines markedly with application of the magnetic field; see Figure 6.

Near the anode, the electric field is radial; only real mobility components carry electric current to the anode. The imaginary component of electron mobility is large, causing electrons to approach the anode in long, slowly-shrinking spirals. During such approach, they produce the ions necessary to the near-the-anode portion of the discharge, and also cause the luminous portion of the discharge to envelop the anode completely, as illustrated in Figure 5.

The Plasma (Positive Column)

Interpretations of positive column phenomena are based on Figures 2, 4, 5, and 6 in the light of the rough relative magnitudes of mobilities represented in Figure 10. No particular claim is made as to correctness of Figure 10 as to absolute values.

THE ZERO MAGNETIC FIELD BASE CONDITION

At zero magnetic field (see Figures 4, 6, and 10) the average longitudinal gradient in the positive column is very gentle, and may be considerably less than the radial gradient. For such a nonmagnetics plasma, the important ion movement in the plasma is radial, in response to a combination of the radial electric gradient and a radial ion concentration gradient. The ions flow outward in all radial directions. The ion and electron concentration within the plasma are about equal, and the plasma is circumferentially enveloped by a positive ion sheath. Recombination occurs only toward the outer edge of and outside the sheath, where random particle energies are small.

The sheath pattern changes when the magnetic field is applied, but a rough equality of plasma electron and ion concentration within the plasma persists, and recombination continues to take place only outside of the plasma proper.

DECLINE OF THE LONGITUDINAL MOBILITIES

Figure 10 indicates that as *B* is increased upward from zero, the first and most rapid change is the decline of the real part of g_e from an initial value around perhaps 100 on the Figure 10 scale to very much smaller values; the real part of g_e rapidly becomes much less than the imaginary part. This initial very rapid trend is real and important, and accounts for the initial rapid increase of the positive column voltage in Figure 6. The plasma gradient becomes larger in order to compel current to flow in spite of the greatly lowered electron mobility.

The increase in gradient caused by the change in g_e results also in a decline in the real part of g_i.[2, 5, 9] Therefore with increasing *B*, the components in the electric field force direction of both mobilities decline markedly. This change in the electron behavior is of predominating importance.

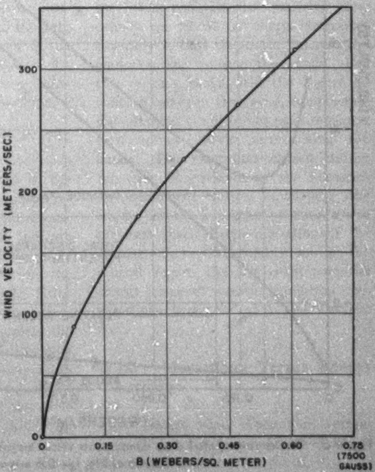


Figure 4. Approximate wind velocity versus magnetic flux density, pressure = 1.2 millimeters of Hg, I_d = 0.45 ampere

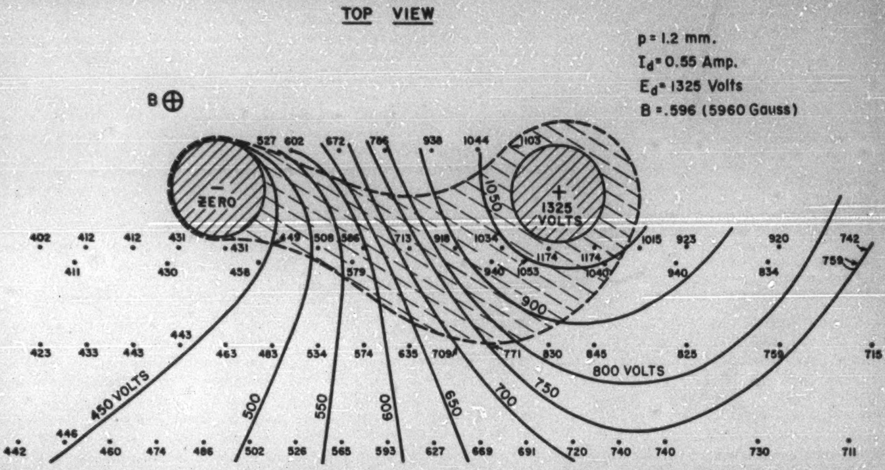

TOP VIEW
p = 1.2 mm.
I_d = 0.55 Amp.
E_d = 1325 Volts
B = .596 (5960 Gauss)
B ⊕
−
ZERO
+
1325 VOLTS
527 602 672 786 938 1044 1103
402 412 412 431 431 449 508 586 713 918 1034 1015 923 920 742
411 430 458 579 940 1053 1174 1174 1040 940 834 759
443
423 433 443 463 483 534 574 635 709 771 830 845 825 759 715
446
442 460 474 486 502 526 565 593 627 669 691 720 740 740 730 711
450 VOLTS
500
550
600
650
700
750
800 VOLTS
900
1050

Figure 5. Potential map of a horizontal plane located at the center (vertically) of the visible discharge. The electrodes are shown in cross section and the approximate visible discharge domain in phantom. Pressure = 1.2 millimeters, magnetic flux density = 5,960 gausses, and discharge current = 0.55 ampere. Potentials are indicated in volts

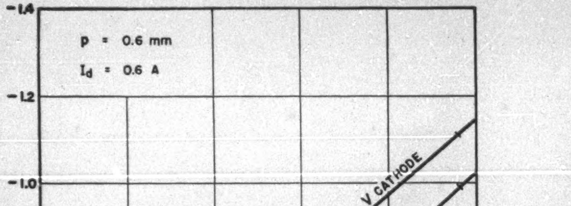


Figure 6. Discharge potential drop components versus magnetic flux density. Pressure = 0.6 millimeter of Hg, I_d = 0.6 ampere

UNBALANCE OF THE POTENTIAL FIELD CAUSED BY A NEGATIVE SPACE CHARGE SHEATH

Figure 5 shows for 6,000 gausses and 1.2 millimeters of mercury a markedly asymmetrical electric field configuration, probably characteristic of all values of B which produce appreciable winds. Study of this potential field, in the light of the Poisson's equation description of the flexion of potential distribution curves, will demonstrate a pronounced negative space charge sheath along the lower boundary only of the luminous (shaded in Figure 5) portion of the discharge. This whole potential field may be pictured as the superposition of a strong and symmetrical longitudinal-gradient field and a field due to a negative space charge sheath on one side only. This space charge exists because, above some modest B, a large electron flow occurs along equipotentials, downward and to the right in Figure 5, because of the predominance of the imaginary component of g_e over the real part. The resulting electron accumulation tends to draw the positive ions downward after it. Poisson's law is the mathematical statement of the way the potential field implements this tendency. Of course, there must exist a positive space charge to balance the charge in the negative sheath; the Figure 5 topography suggests that positive space charge predominates within and above the plasma.

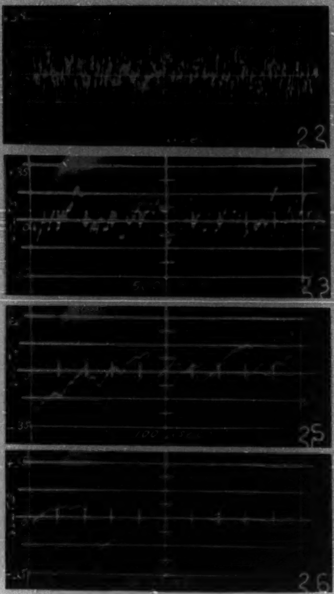


Figure 7 (left). Wave form of potential fluctuations picked up by a floating probe. The horizontal sweep periods of the traces are 5,000, 500, 100, and 50 microseconds, from top to bottom. Pressure = 1.0 millimeter, magnetic flux density = 5,760 gausses, and discharge current = 0.55 ampere

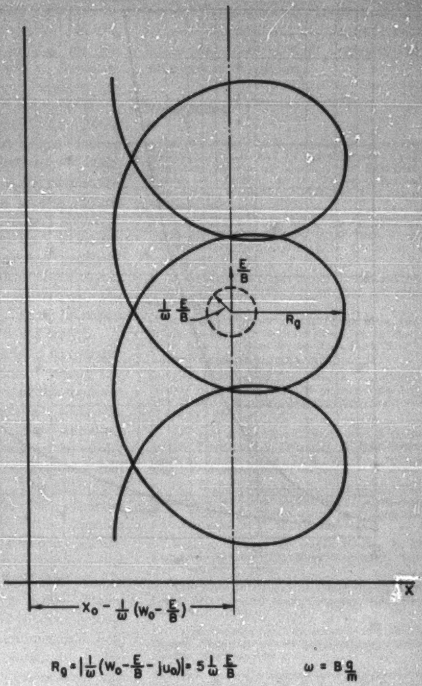

$\frac{E}{B}$
$\frac{1}{\omega}\frac{E}{B}$
R_g
X
$x_o - \frac{1}{\omega}(w_o - \frac{E}{B})$
$R_g = \left|\frac{1}{\omega}(w_o - \frac{E}{B} - ju_o)\right| = 5\frac{1}{\omega}\frac{E}{B}$
$\omega = B\frac{q}{m}$

Figure 8 (right). The x+jz plane projection of the trajectory of an electron with a fairly high initial velocity

Thus the plasma and its environs exhibit the properties of a double line charge, the negative charge existing on the side from which the wind emerges, the positive charge near to and beyond the plasma boundaries on the side into which air enters.

Thus the chain of causative circumstances appears to be as follows:

1. Electron flow occurs chiefly along equipotentials, toward the lower boundary (Figure 5), setting up
2. A negative space charge sheath, causing,
3. The ions to follow the electrons downward because of the transverse component of the electric field due to the negative sheath.

The skewed field shape makes the flow of electrons along equipotentials accomplish part of the total current flow. Thus, because of field asymmetry, the positive column voltage rises less than called for by the decline in real component mobilities alone. This illustrates the general principle that field properties assume whatever co-ordinated pattern is necessary to pass the stated current with a minimum over-all applied voltage.

In the zero-magnetic-field base condition, the equipotential surfaces come close to being cylinders concentric with the axis, because the radial gradient considerably exceeds the longitudinal gradient. Thus quite opposite to one's intuitive concept, the Figure 5 skewed-field equipotentials are probably oriented generally more nearly transverse to the long axis of the discharge than is true in the zero-magnetic-field base condition; the base condition equipotentials are transverse just along the axis, not elsewhere.

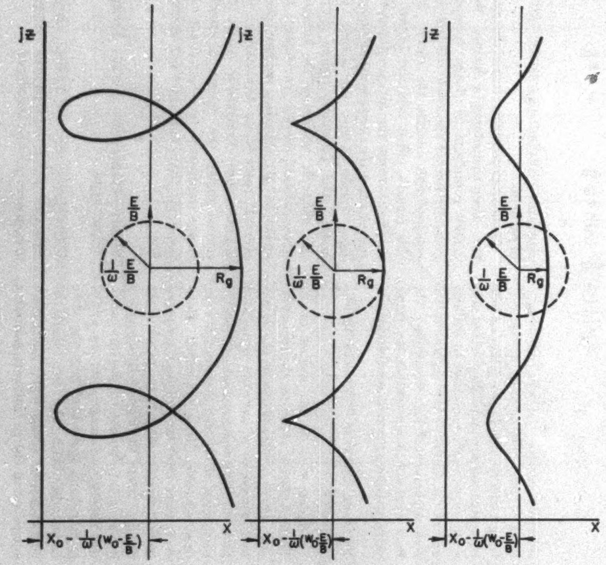
FLOW OF CHARGED PARTICLES IN THE PLASMA

Consider two distinct regions in the Figure 5 plasma. First, for the central region, 450 to perhaps 900 volts, the discussions associated with Figure 10 indicate that the along equipotential (imaginary) component of electron flow provides the most effective charge transport. However, there must be some across-equipotential charge transport to complete the circuit. From the Figure 10 discussion it may well be for the 6,000-gauss condition that the real components of g_e and of g_i may have somewhere near the same magnitude. With the two kinds of charge density about equal, the across-equipotential charge transport may have significant contributions from both kinds of particles.

Second, between 900 volts and the anode the along-equipotential electron flow cannot contribute to charge transport generally toward the anode, because the equipotentials encircle the anode. For this region's radial field the Figure 10 contrast between the two g_e components suggests that the electrons execute many-turn spirals slowly approaching the anode. The excitation caused during this spiralling results in the luminous (cross-hatched) region enveloping the anode, and the ionization produced there prevents space-charge limitation of the electron flow. Presumably as in the central region the across-equipotential charge transport receives significant contributions from both ion and electron flow.

Wind: Momentum Versus Energy Exchange

To discuss wind production, certain basic concepts as to momentum exchange at collisions must be introduced.

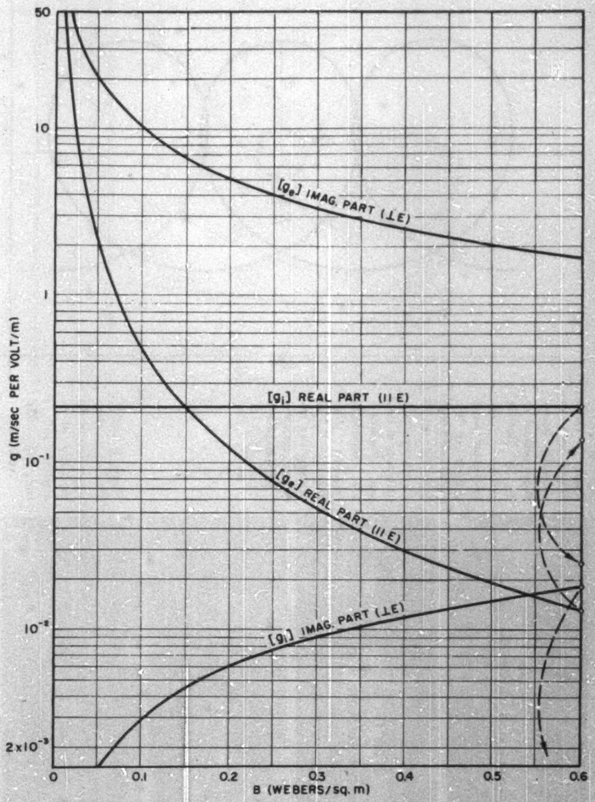


CASE I: $R_g > \frac{1}{\omega}\frac{E}{B}$ CASE II: $R_g = \frac{1}{\omega}\frac{E}{B}$ CASE III: $R_g < \frac{1}{\omega}\frac{E}{B}$

$R_g = \left|\frac{1}{\omega}\left(W_0 - \frac{E}{B} - ju_0\right)\right| \qquad \omega = B\frac{q}{m}$

Figure 9 (above). Projections in the x+jz plane of the types of trajectories typical of positive ions with fairly low initial velocity

Figure 10 (right). Ion and electron mobility versus magnetic flux density. The real components are parallel to the electric field, the imaginary components are perpendicular to the electric field. $\xi_0 = 1.0$ volt for electrons, $\xi_0 = 0.06$ volt for ions. See text discussion of the dashed-line arrows



If a moving ion makes a head-on nonglancing collision with a stationary gas particle of equal mass, the ion comes to rest and the gas particle moves away with just the momentum and the energy brought in by the ion, as in the comparable billiard-ball impact. There is 100-per-cent transfer of momentum and of energy.

If a moving electron makes a head-on nonglancing collision with a stationary gas particle, the electron bounces back with a velocity almost that of approach but oppositely directed. Thus the transfer of momentum is twice that brought into the collision by the electron, whereas the fraction of energy transferred is very minute. There is 200-per-cent transfer of momentum; the fraction of the energy transferred is of the order of the inverse of the mass ratio, about 1/50,000 for air.

As to momentum exchange in a zero-magnetic field plasma:

1. The electron and ion drift velocities are related about inversely with $\sqrt{m_g/m_e}$, where m_g and m_e are gas-particle and electron masses.

2. The momenta (mv) per charged particle will therefore be related about directly as $\sqrt{m_g/m_e}$.

3. For both particles, the momentum transfer to the gas is of the order of unity, that is, 1 or 2; therefore, the momentum transfer per particle at collisions for ions will be about $\sqrt{m_g/m_e}$ times greater than for electrons.

4. The number of collisions per second varies directly with the velocity; therefore the rate of collision for ions will be about $\sqrt{m_g/m_e}$ times less than for electrons.

5. The total momentum transfer per charged particle per second will vary as the product of the transfer per collision (item 3) by the rate of collision occurrence (item 4); therefore each electron will transfer to gas particles about the same momentum per second as each ion.

6. The concentrations of the two kinds of particles are about equal; in a zero-magnetic field plasma they transfer momentum in opposite directions, therefore there is very little net transfer of momentum to the gas and no observable wind.

In a plasma in a strong magnetic field, as for Figure 5, the momentum transfer in the along-axis direction may still be expected to cancel out to zero, but the transverse momentum transfer for the two kinds of particles acts in the same sense, and will therefore combine additively to produce a transverse wind.

In more detail, in Figure 5 the electron drift vector is directed downward and to the right along the skewed equipotentials, because in Figure 10 at 6,000 gausses the g_e imaginary component greatly exceeds the real one. There is not sufficient evidence to tell by how large a factor the g_i real component exceeds the imaginary one; probably the ion drift vector makes some angle between 45 and 90° with the skewed equipotentials, being directed downward and to the left in Figure 5.

However, it appears that at 6,000 gausses the resultant g_e (nearly wholly imaginary) must greatly exeed the resultant g_i, whatever its direction. Thus the g_e/g_i ratio has at 6,000 gausses the same general order of magnitude as the wholly real g_e/g_i ratio in the zero-magnetic field velocity ratio in item 1. Thus just as in that case the small mass of the electrons may be compensated for, as to momentum transfer, by their greater velocity in combination with their greater collision rate caused by their greater velocity. Thus a reasonable although not firmly demonstrated conclusion is that in a strong magnetic field just as in zero-magnetic field the total momentum transfer due to the electrons may be quite comparable with that due to ions. Transverse (wind-producing) momentum transfer due to the ions is certainly important; the point of this argument is to indicate that the transverse momentum transfer due to the electrons may also be important in spite of the very small electron mass.

A significant experimentally observed fact is that when there is a wind, its direction, for any observed values of current, pressure, and magnetic field strength, is at right angles to the plasma boundary lying adjacent to the place of wind measurement. This suggests that there is some general principle present requiring that the angle of skew and the relative magnitudes of the electron and ion mobility components be so related as to cause the resultant momentum transfer to take place at right angles to the main axis of the plasma.

References

1. A Study of the Influence of a Strong Magnetic Field on an Unconfined Glow Discharge in Air at about 1 mm Pressure, **Warren D. McBee.** Ph.D. Thesis (Electrical Engineering), University of Michigan, Ann Arbor, Michigan, 1951.

2. Fundamental Processes of Electrical Discharge in Gases (book), **L. B. Loeb.** John Wiley and Sons, Inc., New York, N. Y., chapter V, part B, 1939.

3. Electricity in Gases (book), **J. S. Townsend.** Oxford University Press, London, England, 1915, pages 96–102.

4. Supersonic Wind at Low Pressures Produced by Arc in Magnetic Field, **H. C. Early, W. G. Dow.** *Physical Review* (New York, N. Y.), volume 79, number 1, July 1, 1950, page 1.

5. Fundamentals of Engineering Electronics (book), **W. G. Dow.** John Wiley and Sons, Inc., New York, N. Y., 1952, second edition, pages 49–52 and 472–87.

6. Studies of Electrical Discharges in Gases at Low Pressures, **L. Langmuir, H. Mott-Smith.** *Physical Review* (New York, N. Y.), volume XXVIII, 1928, page 727.

7. The Characteristics of Electrical Discharges in Magnetic Fields (book), **A. Guthrie, R. K. Wakerling.** McGraw-Hill Book Company, Inc., New York, N. Y., 1949, chapter II.

8. Kinetic Theory of Gases (book), **E. H. Kennard.** McGraw-Hill Book Company, Inc., New York, N. Y., 1938.

9. The Kinetic Theory of Gases (book), **L. B. Loeb.** McGraw-Hill Book Company, Inc., New York, N. Y., 1934.

10. The Mathematic Theory of Non-Uniform Gases, **Chapman, Cowling.** Cambridge Press, London, England, 1939.

11. Preliminary Research on a Low-Pressure Ionic Wind Tunnel, **H. C. Early, W. G. Dow.** *External Memorandum Number 41, Project M-979*, Willow Run Research Center, University of Michigan (Ann Arbor, Mich.), November 1949.

12. Die Beeinflussung der Positiven Saule durch ein Magnetisches Querfeld, **V. T. Wasserrab.** *Zeitschrift fur Naturforschung* (Wiesbaden, Germany), volume IIA, 1947, page 575.

13. Theory of Magnetic Effects in the Plasma of an Arc, **L. Tonks.** *Physical Review* (New York, N. Y.), volume LVI, 1939, page 360.

14. Plasma Electron Drift in a Magnetic Field with a Velocity Distribution Function, **L. Tonks, W. C. Allis.** *Physical Review* (New York, N. Y.), volume LII, 1937, page 710.

15. Arc Motion Reversal in Transverse Magnetic Field, **C. G. Smith.** *Physical Review* (New York, N. Y.), volume LXXIII, 1948, page 548.

16. The Quantitative Study of the Collisions of Electrons with Atoms, **R. B. Brode.** *Reviews of Modern Physics* (New York, N. Y.), volume V, October 1933, pages 257–78 (particularly Figure 11, page 264).

◆

Discussion

H. C. Early, H. L. Smith (University of Michigan, Ann Arbor, Mich.): This is a very interesting presentation of a subject which may be of considerable engineering significance. An electrical discharge of this type may become useful in connection with low density hypersonic wind tunnels. In such an application, the discharge would function as a "booster" to exert an accelerating force on the air stream, and at the same time add thermal energy to the gas. The heating of the stream would permit higher tunnel velocities to be obtained without difficulties due to condensation.

One very important factor affecting the behavior of the "crossed fields" discharge is the large amount of heat energy which is added to the gas and the resulting thermal expansion. Due to this expansion, the gas leaving the electrode region has a much larger volume and velocity than the gas which enters. Therefore, the gas does not continue as a unidirectional wind or jet, but tends to expand in a hemispherical fashion.

A numerical calculation will illustrate the importance of this thermal effect. An order of magnitude calculation of the gas temperature can be obtained by means of a few simplifying approximations. It appears reasonable to assume that in a 3-kilowatt discharge at least 1 kilowatt of power is expended in heating the air. The other assumptions are as follows:

static pressure = 0.5 millimeter of H_g
cross-sectional area of stream = 15 centimeters squared
velocity of entering air = 100 meters per second
temperature of entering air = 20 degrees centigrade
specific heat of air at constant pressure = 0.28 calorie per gram

If these assumptions can be justified, the air stream reaches a temperature of over 8,000 degrees centigrade and the volume increases by a factor of more than 25.

The curve of wind velocity versus magnetic field strength given in Figure 4 was estimated on the basis of the wind pressure on a mica vane. A meaningful interpretation of wind velocity in terms of pressure on a vane would seem to require more information regarding the temperature and density of the gas than is presently available. Changes in magnetic field strength cause changes, not only in the magnetic force on the conducting plasma but also in the power input and the temperature and density of the air stream. Hence, even the shape of this curve is probably of questionable value.

A quantitative study of this type of discharge requires the development of new methods of instrumentation, in order that such quantities as gas density, ion density, temperature, and velocity can be determined under these unusual conditions. At the present time, a University of Michigan research project sponsored by the Office of Ordnance Research is attempting to develop measurement techniques which will permit a more complete study of this subject than is now possible.

W. G. Dow: In regard to the calculations by H. C. Early and H. L. Smith, leading to the 8,000-degree-centigrade temperature figure, my chief comment would be that their figure of 100 meters per second of velocity of entering air is on the high side, which means that their 8,000-degree-centigrade figure is on the low side.

The question as to the existence or otherwise of such temperatures as they mention in the type of environment described is an important and as yet unanswered basic research question. Because of the low pressure, familiar methods of estimating or measuring temperature have no application in this environment. Adequate mechanisms exist within the discharge for the production of gas temperatures of the order of magnitude of 8,000 degrees centigrade or higher, but we do not yet know how to measure them.

A Mathematical Analysis of a Series Circuit Containing a Nonlinear Capacitor

LOUIS A. PIPES
MEMBER AIEE

Synopsis: This paper presents a mathematical analysis of a general series circuit composed of an inductance and a resistance in series with a nonlinear capacitor. The analysis is based on the assumption that the saturation curve of the dielectric material of the capacitor may be represented by a hyperbolic sine function. A 1-term approximation of the forced steady-state oscillation of the circuit produced by the application of a periodic electromotive force and bias potential is obtained. This one term solution gives the best "least squares" approximation to the amplitude and phase of the steady-state charge and current oscillation of the system and reduces to the usual expression for the steady-state current, if the capacitor is linear. The method of analysis is a general one and is applicable to many problems of forced oscillations of nonlinear dynamical systems, both electric and mechanical. It depends on a procedure by which the "mean square error" involved in the 1-term approximation is minimized. Several special cases are considered, and an estimate of the natural frequency of the free oscillations of the circuit is obtained.

THE CIRCUIT analyzed in this discussion is of fundamental importance in the theory of dielectric amplifiers. Dielectric amplifier circuits are based on the fact that certain dielectrics have nonlinear characteristics. As early as 1912, Debye discovered that when atoms having different electron affinities combined, electric dipoles were formed.[1] By analogy to the nonlinearity of magnetic materials containing iron, the nonlinearity of the dielectric constant caused by the alignment of electric dipoles of atomic ions of certain substances has been termed the ferroelectric effect although there is no iron involved.

Many substances that exhibit the ferroelectric effect have been discovered. Rochelle salt was discovered to be a nonlinear dielectric by Valasek[2] in 1921. The ferroelectric properties of the dihydrogen phosphates and arsenates were studied in 1935 by Busch and Scherrer.[3] Wainer and Salomon[4] discovered the unusual dielectric properties of barium titanate in 1942. The ferroelectric properties of the titanates and some of their applications have been studied by A. von Hippel and his associates at the Massachusetts Institute of Technology.[5]

It is the purpose of this paper to develop a mathematical analysis of a basic series circuit containing as a nonlinear element a capacitor whose dielectric exhibits the ferroelectric effect. The basic circuit is illustrated by Figure 1.

The circuit of Figure 1 is composed of a linear inductor L and a linear resistor R in series with a nonlinear capacitor. The potential impressed upon the circuit consists of a constant potential E_0 in series with a harmonic potential $E_m \sin(\omega t+\theta)$. It is assumed that the potential drop across the nonlinear capacitor is given by $V(q)$, where q is the charge separation of the plates of the capacitor.

Kirchhoff's electromotive force law applied to the circuit of Figure 1 leads to the following differential equation

$$L\ddot{q}+R\dot{q}+V(q)=E_0+E_m \sin(\omega t+\theta) \qquad (1)$$

This differential equation is known in the literature as "Duffing's differential equation."[6,7,8] Equations of this kind occur in the mathematical investigations of physical problems of different types, such as the following ones:

1. The forced oscillations of a pendulum due to the application of an applied periodic force.[9]
2. The forced oscillations of a single mass subjected to an elastic restoring force.[10,11,12]
3. The steady-state oscillations of alternating circuits containing iron core inductances.[9]
4. The problem of hunting of synchronous electric machinery.[7,13]

Although an explicit solution of equation 1 in terms of the elementary functions cannot be obtained, several approximation methods of obtaining the solution for various special functions $V(q)$, have been developed.[6,7,8,9]

In the present discussion, a method of obtaining an approximate steady-state or periodic solution of equation 1 will be developed based on a procedure that minimizes the mean of the square of the error introduced by the approximate solution. The method does not depend

---

Paper 53-29, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted July 11, 1952; made available for printing November 14, 1952.

Louis A. Pipes is with the University of California, Los Angeles, Calif.

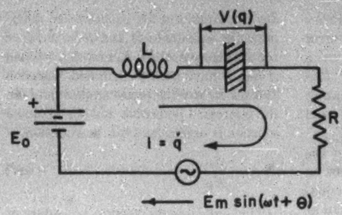


Figure 1 (left). Nonlinear capacitor circuit

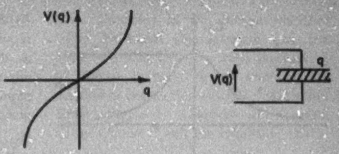


Figure 2 (right). Saturation curve

on the form of the function $V(q)$ but is quite general and

$$L\ddot{q}+G(\dot{q})+V(q)=E_0+E_m \sin(\omega t+\theta) \qquad (2)$$

This equation occurs if the resistance as well as the capacitance of the circuit is nonlinear. In the case of a linear circuit, we have

$$G(\dot{q})=R\dot{q}=Ri \qquad (3)$$

This is the resistive electromotive force drop in the circuit, and

$$V(q)=S_0q=\frac{q}{C_0} \qquad (4)$$

is the capacitive drop. In this case S_0 is is the elastance and C_0 is the capacitance of the linear capacitor.

Saturation Curve of Nonlinear Capacitor

In order to effect a mathematical solution of equation 2 it is necessary to represent the saturation curve $V(q)$ of the nonlinear capacitor by an analytical expression. The saturation curve of a capacitor that has a typical nonlinear dielectric between its plates has the general form given in Figure 2.

The following analytical expression may be adjusted to fit the saturation curves of nonlinear capacitors having typical nonlinear dielectrics.

$$V(q)=\frac{S_0}{a}\sinh(aq) \qquad (5)$$

In this expression, S_0 is the initial elastance of the capacitor defined by the equation

$$S_0=\left(\frac{dV}{dq}\right)_{q=0} \qquad (6)$$

This is the reciprocal if the initial capacitance defined by

$$C_0=\left(\frac{dq}{dV}\right)_{q=0} \qquad (7)$$

The quantity

$$S_i=\left(\frac{dV}{dq}\right)=S_0\cosh(aq) \qquad (8)$$

is the incremental elastance and

$$C_i=\frac{dq}{dV}=\frac{1}{S_0\cosh(aq)}=\frac{C_0}{\cosh(aq)} \qquad (9)$$

is the incremental capacitance.

It may be noted that the quantity a is a measure of the degree of the nonlinearity of the nonlinear capacitor. That is, if $a=0$, we have the linear case, and then $S_i=S_0$, and $C_i=C_0$.

The variation of the incremental capacitance with the charge separation of the capacitor is indicated graphically by Figure 3.

The No-Load Characteristics of the Nonlinear Capacitor

Before considering the general series circuit of Figure 1, it is instructive to consider the case for which $R=0$ and $L=0$. In this case we have a nonlinear capacitor in series with the electromotive force, $E_0+E_m\sin(\omega t)$ since it is not necessary to consider the phase angle, θ because of the absence of the resistance term. Equation 1 now reduces to

$$\frac{S_0}{a}\sinh(aq)=E_0+E_m\sin(\omega t)=E(t) \qquad (10)$$

Hence

$$\sinh(aq)=\frac{aE(t)}{S_0} \qquad (11)$$

and

$$q=\frac{1}{a}\sinh^{-1}\left[\frac{aE(t)}{S_0}\right] \qquad (12)$$

The current in the circuit is given by

$$i=\frac{dq}{dt}=\frac{\left(\frac{dE}{dt}\right)}{\sqrt{a^2E^2(t)S_0^2}} \qquad (13)$$

but

$$\left(\frac{dE}{dt}\right)=\omega E_m\cos(\omega t) \qquad (14)$$

Hence

$$i=\frac{\omega E_m\cos(\omega t)}{\sqrt{a^2[E_0E_m\sin\omega t]^2+S_0^2}} \qquad (15)$$

This is the current in the circuit. If the bias potential E_0 is large compared with the amplitude of the harmonic potential E_m, we have

$$E_0>>E_m \qquad (16)$$

and equation 14 gives the following approximate value for the current

$$i=\frac{\omega E_m\cos(\omega t)}{\sqrt{a^2E^2+S_0^2}} \qquad (17)$$

It may be seen by considering equation 17 that an increase of the bias potential decreases the amplitude of the current and vice versa. The increase of the effective impedance of the nonlinear capacitor with an increase of the bias potential E_0 is the principle on which dielectric amplifiers operate.

The Minimum Mean Squared Error Method for Solving Nonlinear Differential Equations

A method for the approximate solution of a class of nonlinear differential equations in the steady state will now be described. It will be applied to several special cases of the circuit of Figure 1 in later sections.

$$L\ddot{q}+G(\dot{q})+V(q)=E_0+E_1\sin(\omega t)+E_2\cos(\omega t) \qquad (18)$$

where

$$E_1=E_m\cos(\theta) \qquad (19)$$

$$E_2=E_m\sin(\theta) \qquad (20)$$

Hence

$$E_m=\sqrt{E_1^2+E_2^2} \qquad (21)$$

and

$$\tan(\theta)=\frac{E_2}{E_1} \qquad (22)$$

Now from the general theory of differential equations it is known that the equation 18 has solutions, $q(t)$, that are uniquely determined when $q(0)$ and $\dot{q}(0)$ are prescribed, that is, the inital charge separation on the plates of the capacitor and the initial current of the circuit are given and it is clear that solutions of equation 18 that are not periodic exist. Experimental observations of dynamical systems whose equations of motion are of the form of equation 18 indicate that after some transient dis-

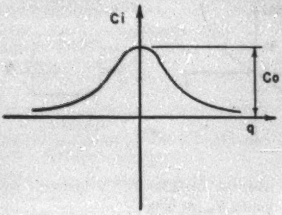


Figure 3. Incremental capacitance

turbances die out the motions of these systems tend to become periodic as the time increases. The experimental and theoretical literature on the subject is devoted almost entirely to the periodic or steady-state solutions.[14,15]

In this analysis, a simple 1-term approximation of the periodic solution of equation 18 will be assumed in the form

$$q_0=A\sin(\omega t)+B=A\sin(T)+B \tag{23}$$

where

$$T=(\omega t) \tag{24}$$

The assumed solution equation 23 may be regarded as the constant and fundamental term of the Fourier series expansion of a general periodic solution whose period is the same as that of the applied harmonic potential.

Let the approximate solution equation 23 be substituted into the left member of equation 18. The resulting equation may be written in the form

$$L\ddot{q}_0+G(\dot{q}_0)+V(q_0)=E_0+E_1\sin(T)+E_2\cos(T)+H(T) \tag{25}$$

Where $H(T)$ is the error introduced because of the approximate nature of the assumed solution q_0.

Let

$$\begin{aligned}F(T)&=L\ddot{q}_0+G(\dot{q}_0)+V(q_0)\\&=-\omega^2LA\sin(T)+G(\omega A\cos T)+V(A\sin T+B)\end{aligned} \tag{26}$$

As a consequence of equations 25 and 26, the error $H(T)$ may be expressed in the form

$$H(T)=F(T)-[E_0+E_1\sin(T)+E_2\cos(T)] \tag{27}$$

The quantity

$$M=\frac{1}{2\pi}\int_0^{2\pi}H^2(T)dT \tag{28}$$

is the mean squared error of the approximate solution taken over a cycle of the oscillation.

It is desired to use equation 27 in some manner to determine the relation between the amplitudes A and B and the constants E_0, E_1 and E_2. To do this, an arbitrary requirement that the mean squared error as defined by equation 28 shall be a minimum will now be imposed. This requirement is suggested by Legendre's principle of "least possible error," and his "method of least squares."[16]

In order to apply the principle of the minimum of the mean squared error to the present situation, the integral, equation 28, will be regarded as a function of E_0, E_1, and E_2. The necessary conditions that M be a minimum are

$$\begin{aligned}\frac{\partial M}{\partial E_0}&=\frac{1}{2\pi}\int_0^{2\pi}\frac{\partial}{\partial E_0}[H^2(T)]dT\\&=-\frac{1}{\pi}\int_0^{2\pi}H(T)dT=0\end{aligned} \tag{29}$$

$$\begin{aligned}\frac{\partial M}{\partial E_1}&=\frac{1}{2\pi}\int_0^{2\pi}\frac{\partial}{\partial E_1}[H^2(T)]dT\\&=-\frac{1}{\pi}\int_0^{2\pi}H(T)\sin(T)dT=0\end{aligned} \tag{30}$$

$$\begin{aligned}\frac{\partial M}{\partial E_2}&=\frac{1}{2\pi}\int_0^{2\pi}\frac{\partial}{\partial E_2}[H^2(T)]dT\\&=-\frac{1}{\pi}\int_0^{2\pi}H(T)\cos(T)dT=0\end{aligned} \tag{31}$$

If the conditions of equations 29, 30 and 31 are used, the following equations are obtained by the integration of equation 27.

$$\int_0^{2\pi}F(T)dT=2\pi E_0 \tag{32}$$

$$\int_0^{2\pi}F(T)\sin(T)dT=\pi E_1 \tag{33}$$

$$\int_0^{2\pi}F(T)\cos(T)dT=\pi E_2 \tag{34}$$

An estimation of the mean squared error may be obtained by a direct integration of equation 28 and the use of equations 32, 33, and 34. After some reductions, the expression may be put into the following form

$$\begin{aligned}M&=\frac{1}{2\pi}\int_0^{2\pi}H^2(T)dT\\&=\frac{1}{2\pi}\int_0^{2}F^2(T)dT-\left[E_0^2+\frac{E_m^2}{2}\right]\end{aligned} \tag{35}$$

This expression is frequently useful in determining the degree of accuracy of the solution by this method.

Equations 32, 33, and 34 give the desired relations between the amplitudes A and B of the assumed solution, equation 23, and the given potential and amplitude E_0 and E_m. Since

$$E_1^2+E_2^2=E_m^2 \tag{36}$$

it can be seen that equations 32 and 36 are two simultaneous equations to determine the unknowns A and B in terms of the known quantities E_0 and E_m. When A and B have been determined, equation 34 fixes E_2 and the phase angle θ between the impressed potential, and the assumed solution is determined by the equation.

$$\sin(\theta)=\frac{E_2}{E_m} \tag{37}$$

This indicates the general procedure of the minimum mean squared error method of solution. The general method will now be applied to the general series circuit under consideration.

Minimum Error Squared Analysis of the General Series Circuit

The differential equation of the general series circuit containing a nonlinear capacitor is

$$L\ddot{q}+R\dot{q}+\frac{S_0}{a}\sinh(aq)=E_0+E_1\sin(T)+E_2\cos(T) \tag{38}$$

For this case, the general function $G(\dot{q})$ and $V(a)$ of equation 25 reduce to

$$G(\dot{q})=R\dot{q} \tag{39}$$

and

$$V(q)=\frac{S_0}{a}\sinh(aq) \tag{40}$$

We also have

$$F(T)=-\omega^2LA\sin(T)+R\omega A\cos(T)+V(q_0) \tag{41}$$

where

$$\begin{aligned}V(q_0)&=\frac{S_0}{a}\sinh(aq_0)=\frac{S_0}{a}\times\sinh[a(A\sin T+B)] \qquad (42)\\&=\frac{S_0}{a}\sinh[P\sin T+Q]\\&=\frac{S_0}{a}\sinh(P\sin T)\cosh(Q)+\frac{S_0}{a}\cosh(P\sin T)\sinh(Q)\end{aligned}$$

and

$$P=aA,\ Q=aB \tag{43}$$

The following identities are established in works on Bessel functions.[17]

$$\sinh(P\sin T)=2\sum_{n=0}^{\infty}(-1)^nI(2n+1)\times(P)\sin(2n+1)T \tag{44}$$

and

$$\cosh(P\sin T)=I_0(P)+2\sum_{n=1}^{\infty}(-1)^nI_{2n}(P)\cos(2nT) \tag{45}$$

The functions $I_n(P)$ are the modified Bessel functions of the first kind of order n and argument P. These functions have been extensively tabulated.[18]

As a consequence of equations 44 and and 45, it can be seen that $V(q_0)$ may be written in the following form

$$V(q_0)=\frac{S_0}{a}I_0(P)\sinh(Q)+\frac{2S_0}{a}I_1(P)\cosh(Q)\sin(T)+\phi(T) \quad (46)$$

where the function $\phi(T)$ contains sines and cosines of multiples of T.

Therefore equation 41 reduces to

$$F(T)=\frac{S_0}{a}I_0(P)\sinh(Q)+\left[\frac{2S_0}{a}I_1(P)\cosh(Q)-\omega^2LA\right]\times \sin(T)+R\omega A\cos(T)+\phi(T) \quad (47)$$

The function $\phi(T)$ has the property that

$$\int_0^{2\pi}\phi(T)dT=0$$
$$\int_0^{2\pi}\phi(T)\sin(T)dT=0 \quad (48)$$
$$\int_0^{2\pi}\phi(T)\cos(T)dT=0$$

as a consequence of the orthogonality property of the trigonometric functions. In this case, equations 32, 33 and 34 reduce to

$$\frac{S_0}{a}I_0(P)\sinh(Q)=E_0 \quad (49)$$

$$\frac{2S_0}{a}I_1(P)\cosh(Q)-\omega^2\frac{LP}{a}=E_1 \quad (50)$$

$$\frac{R\omega P}{a}=E_2 \quad (51)$$

These equations determine P and Q and the phase angle θ. From P and Q, A, and B may be determined by the use of equation 43.

CALCULATION OF AMPLITUDES OF STEADY-STATE OSCILLATION

If equations 50 and 51 are squared and added together, the result may be written in the following form

$$[2S_0I_1(P)\cosh(Q)-\omega^2LP]^2+R^2\omega^2P^2=a^2E_m^2 \quad (52)$$

or

$$2S_0I_1(P)\cosh(Q)=\omega^2LP+\sqrt{a^2E_m^2-R^2\omega^2P^2} \quad (53)$$

From equation 49 we have

$$\sinh(Q)=\frac{aE_0}{S_0I_0(P)} \quad (54)$$

By the use of hyperbolic trigonometry we obtain

$$\cosh(Q)=\sqrt{1-\sinh^2(Q)}=\sqrt{1\frac{a^2E_0^2}{S_0^2I_0^2(P)}} \quad (55)$$

If this value of cosh (Q) is substituted into equation 53, the resulting equation is

$$\frac{2I_1(P)}{I_0(P)}\sqrt{S_0^2I_0^2(P)+a^2E_0^2}=\omega^2LP+\sqrt{a^2E_m^2-R^2\omega^2P^2} \quad (56)$$

This is a transcendental equation for P. If numerical values of the parameters of the circuit are available, it may be solved graphically by plotting the left and right members as functions of P and determining the point of intersection of the two curves obtained in this manner.

After the value of P has been thus determined, equation 53 may be used to determine Q and the result written as

$$Q=\sinh^{-1}\frac{[aE_0]}{S_0I_0(P)} \quad (57)$$

As a consequence of equation 20, the required phase angle is given by

$$\sin(\theta)=\frac{E_2}{E_m}=\frac{R\omega P}{aE_m} \quad (58)$$

Therefore when P is known, θ may be determined by equation 58. In this manner the steady-state oscillation of the charge q, of the general circuit is given by

$$q_0=\frac{P}{\omega}\sin(\omega t)+\frac{Q}{a} \quad (59)$$

and the steady-state current is

$$i_0=\dot{q}_0=\frac{P\omega}{a}\cos(\omega t) \quad (60)$$

as a result of the applied potential

$$E(t)=E_0+E_m\sin(\omega t+\theta) \quad (61)$$

The phase angle θ between the applied harmonic potential and the oscillation of the charge is

$$\theta=\sin^{-1}\frac{[R\omega P]}{aE_m} \quad (62)$$

This outlines the general method of obtaining the approximate steady-state response of the system by the minimum error squared method of analysis. Several special cases of practical importance will now be discussed.

Case of Negligible Resistance and Inductance and $E_0>>E_m$

The case in which the circuit has no resistance or inductance and a large bias potential in comparison with the amplitude of the harmonic potential has been discussed in the section entitled "The No-Load Characteristics of the Nonlinear Capacitor." It is instructive to discuss it as a special case of the method of the section entitled "Minimum Error Squared Analysis of General Series Circuit" with the condition that

$$R=L=0 \quad (63)$$

It will also be assumed that the bias potential E_0 is considerably greater than the maximum value of the applied harmonic potential E_m.

As a consequence of equation 63, equation 56 reduces to

$$\frac{2I_1(P)}{I_0(P)}\sqrt{S_0^2I_0^2(P)+a^2E_0^2}=aE_m \quad (64)$$

The modified Bessel functions $I_0(P)$ and $I_1(P)$ have the following power series expansions.[17]

$$I_0(P)=1+\frac{P^2}{2^2}+\frac{P^4}{2^2\cdot4^2}+\frac{P^6}{2^2\cdot4^2\cdot6^2}+\dots \quad (65)$$

$$I_1(P)=\frac{P}{2}\left(1+\frac{P^2}{2^2\cdot2}+\frac{P^4}{2\cdot2^4\cdot2\cdot3}+\dots\right) \quad (66)$$

In this case equation 50 reduces to

$$2S_0I_1(P)\cosh(Q)=aE_1 \quad (67)$$

Now since E_m is assumed small, therefore aE_1 is also small and because cosh(Q) is never less than 1, it follows that $I_1(P)$ is small and hence may be replaced by the first term of its power series expansion, equation 66. Hence we have in this case

$$I_1(P)\doteq\frac{P}{2} \quad (68)$$

and

$$I_0(P)\doteq1 \quad (69)$$

Therefore equation 64 reduces to

$$P\sqrt{(S_0^2+a^2E_0^2)}=aE_m \quad (70)$$

or

$$P=\frac{aE_m}{\sqrt{S_0^2+a^2E_0^2}} \quad (71)$$

and as a consequence of equation 57 we have

$$Q=\sinh^{-1}\frac{[aE_0]}{S_0} \quad (72)$$

By equation 60 the steady-state current is

$$i_0=\frac{\omega E_m\cos(\omega t)}{\sqrt{S_0^2+a^2E^2}} \quad (73)$$

This result agrees with equation 16.

General Case with Large Bias Potential

The general case for which the condition

$$E_0 >> E_m \tag{74}$$

is satisfied will now be considered because of its practical importance. Since the assumed value of E_m is small, it follows that aE_1 and aE_2 are small quantities and hence as a consequence of equations 50 and 51 the value that P can take is so small that the Bessel functions, $I_0(P)$ and $I_1(P)$ may be replaced by the first terms of their Mclaurin series expansions equations 65 and 66 as given by equations 68 and 69. Hence equation 52 now takes the form

$$[PS_0 \cosh(Q) - \omega^2 LP]^2 + R^2\omega^2P^2 = a^2E_m^2 \tag{75}$$

or

$$P = \frac{aE_m}{\sqrt{R^2\omega^2 + [S_0 \cosh(Q) - \omega^2 L]^2}} \tag{76}$$

and equation 49 now takes the form

$$\sinh(Q) = \frac{aE_0}{S_0} \tag{77}$$

Hence

$$S_0 \cosh(Q) = \sqrt{S_0^2 + a^2E_0^2} \tag{78}$$

If equations 76, 77, and 78 are substituted into equation 59, the following result is obtained.

$$q_0 = \frac{E_m \sin(\omega t)}{\sqrt{R^2\omega^2 + \sqrt{[S_0^2 a^2E_0^2]} - \omega^2 L)^2}} + \frac{1}{a} \sinh^{-1} \frac{[aE_0]}{S_0} \tag{79}$$

The phase angle is given by equation 58 in the form

$$\sin(\theta) = \frac{R\omega}{\sqrt{R^2\omega^2 + [\sqrt{(S_0^2 + a^2E_0^2)} - \omega^2 L]^2}} \tag{80}$$

If $a=0$, equation 79 reduces to the usual linear solution for steady-state charge oscillation

$$q_0 = \frac{E_m \sin(\omega t)}{\sqrt{R^2\omega^2 + (S_0^2 - \omega^2 L)^2}} + \frac{E_0}{S_0} \tag{81}$$

The steady-state current oscillation of the circuit may be written in the form

$$i_0 = \dot{q}_0 = \frac{E_m \cos(\omega t)}{\sqrt{R^2 + \left[\omega L - \frac{1}{\omega}\sqrt{(S_0^2 + a^2E_0^2)}\right]^2}} \tag{82}$$

The effect of the bias potential in increasing the effective impedance of the circuit is clearly evident.

Natural Frequency of Free Oscillations

Equation 50 may be used to determine the natural frequency of the free oscillations of the circuit in the absence of applied electromotive forces and resistance. To study this case we have

$$E_0 = E_m = R = 0 \tag{83}$$

Hence $E_1 = E_2 = 0$ and $Q = 0$ by equation 49. Equation 50 now reduces to

$$2S_0 I_1(P) = \omega^2 LP \tag{84}$$

Hence

$$\omega^2 = \frac{2S_0}{LP} I_1(P) \tag{85}$$

This equation determines the angular frequency ω of the free oscillations of the circuit in the form

$$q = A \sin(\omega t) \tag{86}$$

Since the amplitude A is related to P by $aA = P$, equation 85 may be written in terms of the amplitude, in the following manner.

$$\omega^2 = \frac{2S_0 I_1(aA)}{LaA} \tag{87}$$

In the linear case, $a=0$ and equation 87 becomes indeterminate. The indeterminate form may be evaluated by taking $I_1(aA) = aA/2$ by equation 66 and we have

$$\omega^2 = \frac{S_0}{L}, \text{ or } \omega = \sqrt{\frac{S_0}{L}} = \sqrt{\frac{1}{LC_0}} \tag{88}$$

for this case.

In the general nonlinear case however, the natural frequency is a function of the amplitude A and is given by equation 87 when the amplitude is known.

Conclusion

The analysis presented in this paper is an attempt to apply the method of "the least mean square error" to the solution of the basic equation of the general series circuit containing a nonlinear capacitor. Although the analysis is based on the representation of the saturation curve of the nonlinear capacitor $V(q)$ by a hyperbolic function, this is not necessary. The general procedure of the method as described in one section entitled "Minimum Error Squared Analysis of the General Series Circuit" may be applied to functions $V(q)$ represented by other analytical expressions. The least mean square error method may be used to obtain approximate expressions for the harmonic steady-state oscillations of many nonlinear electric circuits of practical importance.

References

1. Polar Molecules (book), **P. Debye.** Dover Publications, New York, N. Y., reprinted 1950.
2. J. Valasek. *Physical Review* (New York, N. Y.), volume 17, 1921, page 475.
3. Neue Seignette-Elektrika, **G. Busch.** *Acta Physica Austriaca* (Vienna, Austria), volume 11, 1938, pages 269–298.
4. High Titania Dielectrics, **E. Wainer.** *Transactions*, The Electrochemical Society (New York, N. Y.), volume 89, 1946.
5. Dielectrics Made to Order, **A. von Hippel.** *Electronics* (New York, N. Y.), volume 24, number 3, 1951.
6. Erzwungene Schwingungen bei veränderlicher Eigenfrequenz, **G. Duffing.** F. Vieweg u. Sohn (Braunschweig, Germany), 1918.
7. Nonlinear Vibrations, **J. J. Stoker.** Interscience (New York, N. Y.), 1950, pages 81–117.
8. Ordinary Nonlinear Differential Equations in Engineering, **N. W. McLachlan.** Oxford University Press (London, England), 1950, pages 57–66.
9. Operational Methods in Nonlinear Mechanics, **L. A. Pipes.** *Report 51-10*, Department of Engineering, University of California (Los Angeles, Calif.), 1951.
10. On the Forced Vibrations of Nonlinear Springs, **C. R. Wylie, Jr.** *Journal*, Franklin Institute (Philadelphia, Pa.), volume 236, 1943, page 273.
11. An Experimental Investigation of Forced Vibrations in a Mechanical System Having Nonlinear Restoring Force, **C. A. Ludeke.** *Journal of Applied Physics* (New York, N. Y.), volume 17, 1946, pages 603–09.
12. Forced Vibrations of Systems with Nonlinear Restoring Force, **K. O. Friedrichs, J. J. Stoker.** *Quarterly of Applied Mathematics*, Brown University (Providence, R. I.), volume 1, 1943, pages 97–115.
13. Transient Torque-Angle Characteristics of Synchronous Machines, **W. V. Lyon, H. E. Edgerton.** *AIEE Transactions*, volume 49, April 1930, pages 686–98.
14. Lectures on Differential Equations, **Solomon Lefschetz.** Princeton University Press (Princeton, N. J.), 1948, pages 204–09.
15. Existence of Periodic Solutions for Second Order Differential Equations with a Forcing Term, **N. Levinson.** *Journal of Mathematics and Physics* (Cambridge, Mass.), volume 22, 1943, pages 41–48.
16. The Calculus of Observations (book), **F. T. Whittaker, G. Robinson.** Blackie and Son, Ltd., London, England, 1924, pages 209–59.
17. Bessel Functions for Engineers (book), **N. W. McLachlan.** Oxford University Press, London, England, 1934, page 50.
18. Tables of Functions (book), **E. Jahnke, F. Emde.** Dover Publications, New York, N. Y., 1934, pages 232–33.

◆

Discussion

Thomas M. Stout (University of Washington, Seattle, Wash.): This paper, possibly the first to analyze a resonant circuit containing a nonlinear capacitor, is an interesting and useful addition to the general literature on nonlinear dynamic systems. One implication of the paper, however, seems to merit some discussion.

Professor Pipes states that his equation 1 is "Duffing's equation" and cites a number of references in which this equation is discussed. The uninitiated reader might easily get the idea that the equation considered in this paper has already been solved by other methods and that solutions to this circuit problem could have been obtained by suitable interpretation of existing solutions. Study of several of the references indicates that this is probably not the case.

For example, Stoker[1] first discusses (page 83) the equation

$$\ddot{x}+c\dot{x}+(\alpha x+\beta x^3)=F\cos\omega t \qquad (1)$$

which he describes as Duffing's equation. Later (page 112) he considers the equation

$$\ddot{x}+\alpha x-\beta x^3=H_1\cos\omega_1 t+H_2\cos\omega_2 t \qquad (2)$$

and tabulates (pages 116–17) solutions of these equations for various special cases.

McLachlan[2] discusses the two equations

$$\ddot{y}+ay+by^3=f\cos\omega t \qquad (3)$$

$$\ddot{y}+2k\dot{y}+ay+by^3=f\cos(\omega t+\phi) \qquad (4)$$

Lyon and Edgerton[3] consider the equation

$$P_j\ddot{\theta}+P_d\dot{\theta}+P_m\sin\theta=P_0+P_L \qquad (5)$$

in which P_0 is the initial load on the synchronous machine and P_L is a suddenly applied constant load.

These investigators, at least, do not consider a forcing function which is a constant plus a sinusoid, as Pipes does. Because solutions of nonlinear differential equations cannot be superimposed, any solutions for these other forcing functions are not very useful in the present problem. The analysis which Pipes presents is evidently not an alternative procedure for obtaining known solutions, but a treatment of a previously unsolved problem.

This impression is strengthened by attempts to use his equation 1 in analyzing a series resonant circuit containing a nonlinear inductor. Both Stoker[1] and Minorsky[4] indicate that the differential equation of this circuit might be written as

$$\ddot{\phi}+\frac{1}{C}(\alpha\phi-\beta\phi^3)=0 \qquad (6)$$

where ϕ is the magnetic flux, if the circuit contains no resistance and no voltage sources; this is presumably the sort of equation referred to by Pipes.[5] The second restriction is not essential, but including the resistive term leads to an equation similar to equation 2 of the paper, as Keller shows,[6] not to equation 1. While the method described in this paper would be applicable for Keller, the solutions are not interchangeable.

Many investigations of circuits containing nonlinear inductors have been made, and much labor could be saved if the results of this work could be used to explain the properties of circuits containing nonlinear capacitors. Study of this attractive possibility has led to consideration of the principle of duality as it applies to nonlinear systems.[7]

As in the case of linear networks, the dual of a nonlinear network is obtained by replacing voltages by currents, currents by voltages, inductance by capacitance, resistance by conductance, capacitance by inductance, and series circuits by parallel circuits. A nonlinear inductor becomes a nonlinear capacitor, and conversely. Two examples of series nonlinear circuits and their duals are given in Figure 1 of this discussion.

For complete correspondence between the original network and its dual, the nature of the nonlinearity should be such that one analytic approximation can be used. Surprisingly or naturally enough, depending on the viewpoint, it happens that this is possible for the ordinary nonlinear dielectric and magnetic materials. The hyperbolic sine employed by Pipes is a permissible, though not entirely satisfactory, approximation to the magnetization curve for some materials and has been used in work on magnetic amplifiers.[8-12]

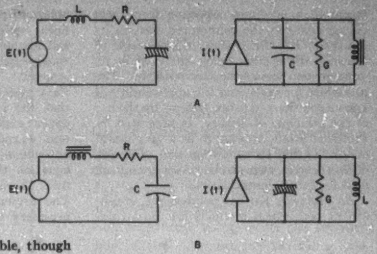


Figure 1. Two series resonant circuits containing a nonlinear element and their duals

Following the pattern of equation 5 of the paper, we can write that the current in an iron-core inductor is

$$i_L=\frac{1}{bL_0}\sinh b\lambda \qquad (7)$$

where b is a parameter used in curve fitting, L_0 is the initial inductance, and $\lambda=N\phi$, the flux linkages. In the linear case, $b=0$ and $L=L_0$.

If this approximation can be used, the circuits shown in Figure 1 of the discussion are complete duals and solutions can be interchanged merely by changing the notation. Consider the circuit treated by Pipes, Figure 1(A) of the discussion. The equations for this circuit are

$$e_c=E(t)-Ri-L\frac{di}{dt} \qquad (8)$$

$$=\frac{1}{aC_0}\sinh aq \qquad (9)$$

$$q=\int i\,dt \qquad (10)$$

For the parallel circuit which is the dual of this series circuit, the corresponding equations are

$$i_L=I(t)-Ge-C\frac{de}{dt} \qquad (11)$$

$$=\frac{1}{bL_0}\sinh b\lambda \qquad (7)$$

$$\lambda=\int e\,dt \qquad (12)$$

The duality of the other pair of circuits can be established in similar fashion.

Professor Pipes has made an important contribution to the collection of solutions available for series resonant circuits containing a nonlinear element. Solutions for the dual parallel circuits can be obtained with very little additional labor.

REFERENCES

1. See reference 7 of paper.

2. See reference 8 of paper.

3. See reference 13 of paper.

4. INTRODUCTION TO NONLINEAR MECHANICS (book), **N. Minorsky.** J. W. Edwards, Ann Arbor, Mich., 1947, page 192.

5. See reference 9 of paper.

6. MATHEMATICS OF MODERN ENGINEERING, VOLUME II (book), **E. G. Keller.** John Wiley and Sons, Inc., New York, N. Y., 1945, pages 275–76.

7. SOME GENERAL THEOREMS FOR NONLINEAR SYSTEMS POSSESSING REACTANCE, **C. Cherry.** *Philosophical Magazine* (London, England), volume 42, October 1951, pages 1161–77.

8. THE PROBLEM OF THE MAGNETIC AMPLIFIER AND SOME APPROACHES TO ITS SOLUTION, **M. Liwschitz-Garik, E. J. Smith, E. Weber.** *Proceedings, National Electronics Conference* (Chicago, Ill.), volume V, 1949, pages 235–54.

9. ANALYTICAL DETERMINATION OF CHARACTERISTICS OF MAGNETIC AMPLIFIERS WITH FEEDBACK, **D. W. Ver Planck, L. A. Finzi, D. C. Beaumariage.** *AIEE Transactions*, volume 68, part I, 1949, pages 565–70.

10. AN ANALYSIS OF TRANSIENTS AND FEEDBACK IN MAGNETIC AMPLIFIERS, **Walter C. Johnson, Frank W. Latson.** *AIEE Transactions*, volume 69, part I, 1950, pages 604–11.

11. GENERAL CHARACTERISTICS OF MAGNETIC AMPLIFIERS, **L. A. Finzi, D. C. Beaumariage.** *AIEE Transactions*, volume 69, part II, 1950, pages 919–27.

12. THE TRANSIENT RESPONSE OF MAGNETIC AMPLIFIERS—CASES OF NEGLIGIBLE COMMUTATION, **L. A. Finzi, D. P. Chandler, D. C. Beaumariage.** *AIEE Transactions*, volume 70, part I, 1951, pages 934–42.

Y. H. Ku (University of Pennsylvania, Philadelphia, Pa.): I should like to congratulate Professor Pipes for his fine contribution to the mathematical analysis of a nonlinear problem. With the recollecting that Dr. Pipes has previously analyzed the nonlinear problem with a variable inductance, it occurs to me that a series circuit with nonlinear inductance and nonlinear capacitance may be further studied by various methods. A series circuit with the elements R, L, C is truly a simple circuit. However, when one, two, or three elements of this series circuit become nonlinear, the solution will be rather complicated. We need more analytical results[1-5] and yet we like to see some of these results expressed in graphical form.[2-5] Besides the isocline method and the Liénard construction, it may be mentioned that another graphical method has been presented.[6]

REFERENCES

1. See reference 6 of Mr. Stout's discussion.

2. See reference 4 of Mr. Stout's discussion.

3. See reference 7 of the paper.

4. See reference 8 of the paper.

5. THEORY OF OSCILLATIONS (book), **A. A. Andronow, C. E. Chaikin.** Princeton University Press, Princeton, N. J., 1949.

6. NONLINEAR ANALYSIS OF ELECTRO-MECHANICAL PROBLEMS, **Y. H. Ku.** *Journal*, Franklin Institute (Philadelphia, Pa.), January 1953.

Louis A. Pipes: Professor Stout emphasizes the fact that solutions of nonlinear differential equations cannot be superimposed and that because of this, addition of terms to a nonlinear equation frequently changes entirely the method required for its solution. For this reason, it is desirable to consider the system being analyzed in its most general form wherever possible. The procedure for extending the applicability of known solutions to other systems by the use of the principle of duality, as suggested by Professor Stout, appears to be a very useful one. The full consideration and extension of this process seems to be a very promising field for future investigation.

As Professor Ku has pointed out, graphical or topological methods are very valuable and powerful in the analysis of nonlinear differential equations. The graphical methods of the phase-plane have been widely used in the analysis of autonomous systems. The differential equations of these systems are characterized by the fact that the independent variable t does not appear explicitly in the equation. The system considered in the paper is a nonautonomous system and the differential equation that governs its behaviour contains the independent variable t explicitly. The presence of the variable t makes it necessary to modify the isocline and the Liénard techniques considerably before they can be adapted to the solution of this problem.

Professor L. S. Jacobsen has given a powerful graphical procedure for the solution of the differential equations of nonautonomous systems if they are of the second order.[1]

REFERENCE

1. ON A GENERAL METHOD OF SOLVING SECOND-ORDER DIFFERENTIAL EQUATIONS BY PHASE-DISPLACEMENTS, L. S. Jacobsen. *Journal of Applied Mechanics* (New York, N. Y.), volume 19, December 1952, page 543.

Influence of A-C Reactance on Voltage Regulation of 6-Phase Rectifiers

R. L. WITZKE
FELLOW AIEE

J. V. KRESSER
MEMBER AIEE

J. K. DILLARD
ASSOCIATE MEMBER AIEE

INVESTIGATIONS made during the past few years have shown that methods normally used for the calculation of rectifier voltage regulation have definite limitations not generally appreciated by application engineers. It has been established that conventional methods of calculating regulation are valid only if the angle of overlap is less than 360 degrees divided by the number of phases. With larger overlap angles any a-c reactance common to two or more simple rectifiers introduces a delay angle, even though the rectifiers are operated without intentional delay. This inherent delay reduces the direct voltage and therefore increases the voltage regulation. In special applications involving large overlap angles, these limitations are of primary importance, because considerable error is possible if they are not recognized.

The purpose of this paper is to discuss the influence of common a-c reactance on the operation of 6-phase rectifiers, to present equations that can be used to calculate regulation in applications where conventional equations are not applicable, and to present curves for estimating rectifier voltage regulation.

Nomenclature

α = angle of forced commutation delay
δ = angle of simultaneous commutation at each end of the commutating period
$\omega = 2\pi f$
$E_{a,b,c}$ = source line-to-neutral instantaneous voltages
$E_{a',b',c'}$ = transformer primary line-to-neutral instantaneous voltages
$E_{1,2,3}$ = instantaneous anode voltages
e_d = instantaneous direct voltage under load
E_d = average direct voltage under load
E_{d_0} = average direct voltage at no-load
E_s = transformer secondary line-to-neutral rms voltage at no load
i = instantaneous value of commutation current
$i_{a,b,c}$ = instantaneous alternating currents in source
I_c = commutated direct current
I_cX_c/E_s = reactance factor based on commutated direct current
I_d = load current
L = commutating inductance
$p = d/dt$
t = time in seconds
u = commutating (overlap) angle
X_c = commutating reactance $= X_s + X_t$
X_s = source reactance only
X_t = reactance in transformer circuit only

Summary

Mathematical expressions have been derived for calculating the voltage regulation of the delta, 6-phase, delta, double-way (full-wave bridge) rectifier for load currents between zero and short-circuit values. Calculated curves are presented for estimating voltage regulation. In addition, regulation curves obtained in the laboratory are included to substantiate the general trend of the calculated curves.

The analysis is based on rectifiers operating without phase control; that is, the direct voltage is not controlled by delaying the firing time of ignitor circuits or by grid control. As is customary, the load circuit is assumed to have sufficient inductance to sustain an essentially flat-topped load current.

Although this paper deals with specific transformer connections, by following a similar analysis it can be shown that the equations derived are not affected by transformer connection. The most common basic rectifier power circuit is the 6-phase double-wye. Since double-wye rectifier transformers are constructed with the secondary windings closely linked together, the reactance between secondary windings is small compared to the leakage reactance between the primary and either secondary winding.[1] If the leakage reactance between secondaries is zero, the double-wye primary circuit reduces to the same circuit as that of the full-wave bridge. Therefore, for practical purposes the curves and equations presented in this paper are valid for double-wye circuits.

In this paper it is shown that there are three modes of operation, as load current is varied from zero to short-circuit value. In the first mode of operation, as load current is increased from no load, the commutating or overlap angle u increases from zero to 60 degrees. During the second mode, the commutating angle is fixed at 60 degrees, but the start of commutation is delayed by an inherent delay angle α, which varies between zero and 30 degrees as the rectifier supplies increasing load current. The angle of inherent delay does not increase during the third mode, while the commutating angle increases to 120 degrees at short circuit. The three modes of operation are summarized in Table I, where I_cX_c/E_s = reactance factor based on commutated direct current.

In the full-wave bridge circuit the commutated direct current I_c is the same as the load current I_d. In the double-wye

---

Paper 53-54, recommended by the AIEE Electronic Power Converters Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted October 22, 1952; made available for printing November 20, 1952.

R. L. WITZKE and J. K. DILLARD are with the Westinghouse Electric Corporation, East Pittsburgh, Pa.; J. V. KRESSER is with the Westinghouse Electric Corporation, San Francisco, Calif.

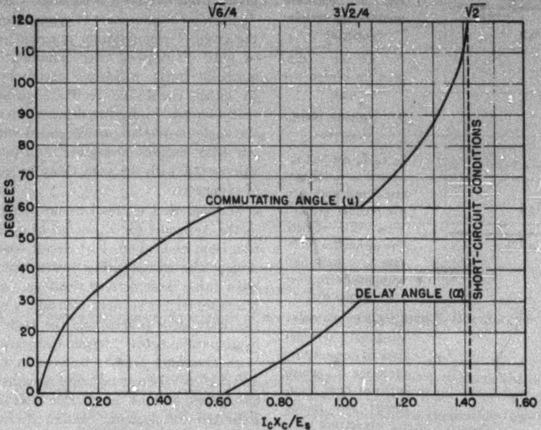


Fig. 1. Commutating angle and forced delay angle in 6-phase bridge rectifier

rectifier the commutated direct current is one-half of the load current.

Calculated values of commutating angle u and angle of inherent delay α are plotted against reactance factors in Fig. 1. The ratio of direct voltage to no-load direct voltage (E_d/E_{do}) is plotted against reactance factors in Fig. 2. Reactance factors may be converted to per-unit reactance drop by multiplying by $2/\sqrt{6}$. This figure is obtained by expressing I_cX_c in per unit on the transformer primary kilovolt-ampere base. Equations used in calculating the curves are derived in the appendix and for convenience are summarized now.

Mode 1,

$$u \leqq 60°, \quad \alpha = 0, \quad \frac{I_cX_c}{E_s} \leqq \frac{\sqrt{6}}{4}$$

$$\frac{I_cX_c}{E_s} = \frac{\sqrt{6}}{2}(1-\cos u)$$

$$\frac{E_d}{E_{do}} = 1 - \frac{I_cX_c}{\sqrt{6}E_s}$$

Mode 2,

$$u = 60°, \quad \alpha \leqq 30°, \quad \frac{\sqrt{6}}{4} \leqq \frac{I_cX_c}{E_s} \leqq \frac{3\sqrt{2}}{4}$$

$$\frac{I_cX_c}{E_s} = \frac{\sqrt{6}}{2}\sin(\alpha+30°)$$

$$\frac{E_d}{E_{do}} = \frac{\sqrt{3}}{2}\cos(\alpha+30°)$$

Mode 3,

$$60° \leqq u \leqq 120°, \quad \alpha = 30°,$$
$$\frac{3\sqrt{2}}{4} \leqq \frac{I_cX_c}{E_s} \leqq \sqrt{2}$$

$$\frac{E_d}{E_{do}} = \frac{\sqrt{3}}{2}[1-\sin(u-30°)]$$

$$\frac{I_cX_c}{E_s} = \frac{\sqrt{2}}{2} + \frac{\sqrt{6}}{4}\sin u - \frac{\sqrt{2}}{4}\cos u$$

Table I. Three Modes of Operation

Mode	u	α	I_cX_c/E_s
1	0→60	0	0→$\sqrt{6}/4$
2	60	0→30	$\sqrt{6}/4$→$(3\sqrt{2})/4$
3	60→120	30	$(3\sqrt{2})/4$→$\sqrt{2}$

Qualitative Discussion

A 3-phase double-way rectifier, commonly referred to as a 6-phase bridge or a full-wave bridge rectifier, is shown in Fig. 3. It consists of two simple rectifiers, group I and group II, in series across the d-c bus. The total reactance to the rectifying elements, including transformer and source reactance, which in this circuit are directly additive when on the same base, is designated as the commutating reactance X_c. In the discussion to follow, the line-to-neutral voltages (E_a, E_b, and E_c) back of the commutating reactance are assumed to have *a-b-c* phase rotation. Line-to-neutral voltages on the rectifier side of the commutating reactance are designated as E_1, E_2, and E_3 for simple rectifier group I and $-E_1$, $-E_2$, and $-E_3$ for simple rectifier group II.

When the rectifier operates without phase control, that is, without delay, the tubes, or rectifying elements, T_1, T_2, T_3, T_4, T_5, and T_6, pick up or start to conduct in the sequence defined by the subscripts. These conductions start at times t_1, t_2, t_3, etc., as shown in Fig. 4. Normally, in calculating rectifier performance the line-to-neutral voltages on the two sides of the commutating reactance are assumed to be equal, that is, $E_1 = E_a$, $E_2 = E_b$, etc., as shown in Fig. 4. In many applications this assumption results in valid conclusions because the assumed relations do exist at the beginning of each conduction period. However, in many other applications this assumption is not valid and can result in considerable error. To illustrate the latter contention, consider the time interval following the start of conduction in tube T_1. During this period tubes T_4, T_5, and T_6 are conducting, and terminals 1 and 3 are short-circuited through T_1 and T_5, resulting in a system short-circuit current i flowing as illustrated in Fig. 3. From circuit theory the following relations are established

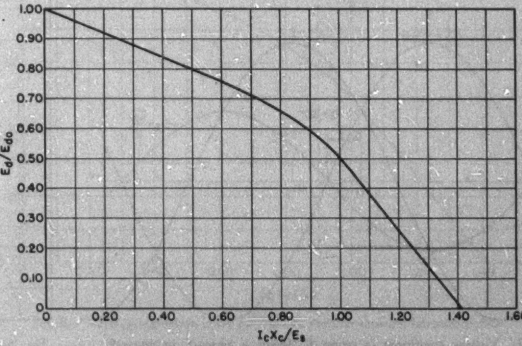


Fig. 2. Voltage regulation curve for 6-phase rectifier

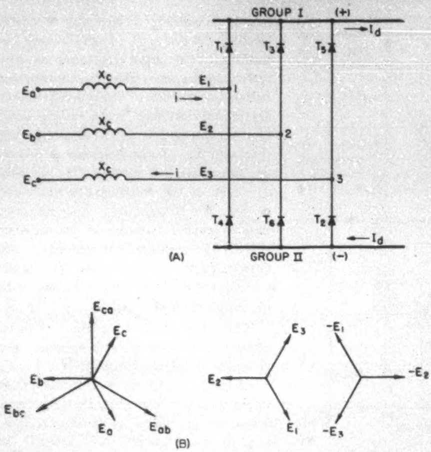


Fig. 3. Six-phase double-way rectifier circuit

A: Circuit diagram for bridge circuit

B: Voltage phasor diagrams

E_1, E_2, E_3 are line-to-neutral voltages associated with simple rectifier group I $-E_1$, $-E_3$, $-E_2$ are line-to-neutral voltages associated with simple rectifier group II

E_a, E_b, E_c are line-to-neutral voltages back of commutating reactance X_c, which includes both transformer and source reactance

$$i=\frac{E_a-E_c}{j2X_c}$$

$$-E_3=-(E_c+jiX_c)=\frac{E_b}{2}$$

$$-E_2=-E_b$$

Sixty degrees after tube T_1 picks up, tube T_2 should start to conduct; however, this conduction cannot start until voltages $-E_3$ and $-E_2$ are equal. Reference to Fig. 5 will show that these voltages are not equal until 30 degrees after time t_2, the instant of a normal, undelayed commutation, if the T_5 to T_1 commutation in group I has not been completed. Tube T_2 therefore cannot start to conduct until the group I commutation is completed, or until 30 degrees after commutation normally starts, whichever occurs first. In either event the rectifier operates with a forced, or inherent, delay whenever the angle of overlap, as calculated without delay, exceeds 60 degrees. All tubes will operate with the same delay because of symmetry.

Mathematical expressions are derived in the Appendix for all modes of operation from no-load to short-circuit conditions. The curves in Fig. 1 were calculated using these derived relations. For reactance factors between zero and $\sqrt{6}/4$ the delay angle is zero, and the angle of overlap varies between zero and 60 degrees. With reactance factors between $\sqrt{6}/4$ and $(3\sqrt{2})/4$ the overlap angle remains at 60 degrees, and the delay angle increases from zero to 30 degrees. Within this zone the rectifier operates with sufficient delay to commutate the load current within 60 degrees. One commutation starts just when the previous one ceases, no simultaneous commutations taking place. When the reactance factor is larger than $(3\sqrt{2})/4$, one commutation starts before the previous one ceases, and so simultaneous commutations occur. The delay

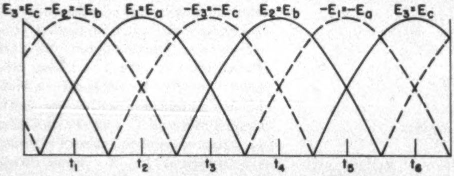


Fig. 4. Voltage relations normally assumed in calculating rectifier performance

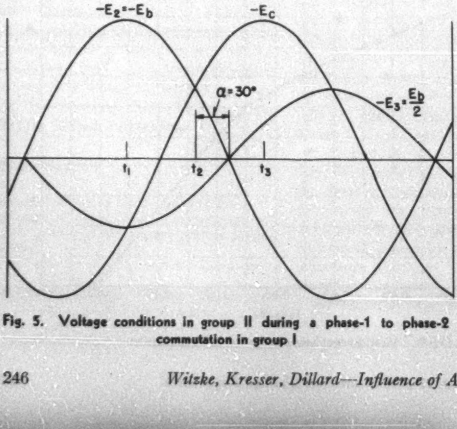


Fig. 5. Voltage conditions in group II during a phase-1 to phase-2 commutation in group I

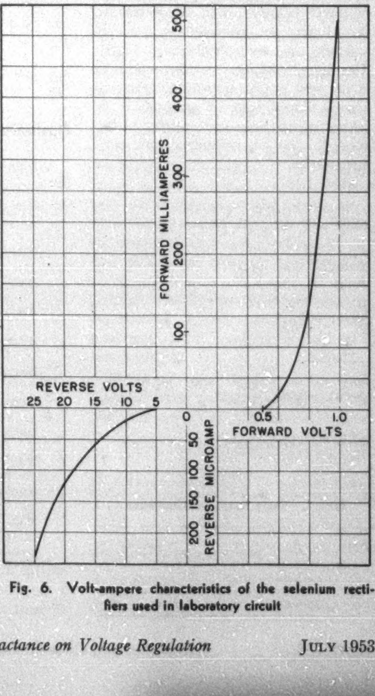


Fig. 6. Volt-ampere characteristics of the selenium rectifiers used in laboratory circuit

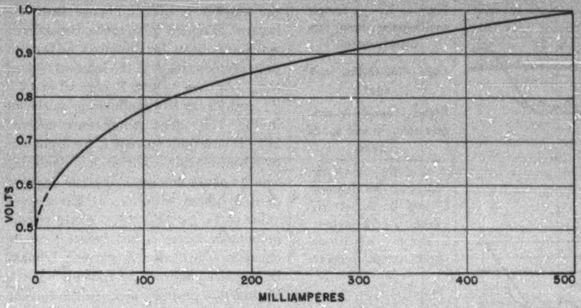


Fig. 7. **Forward volt-ampere characteristic of the selenium rectifiers used in laboratory circuit**

angle remains at 30 degrees, and the overlap angle varies between 60 and 120 degrees, depending upon the reactance factor, which is $\sqrt{2}$ at short circuit.

It has been recognized previously that rectifiers pass through various modes of operation between no load and short circuit,[2,3] and the first two modes outlined here agree with previous work. The contention that the overlap angle increases beyond 60 degrees in the third mode differs from prior analyses.

Laboratory Studies

The analog computer, or Anacom, is a useful tool for studying complicated circuits where analytical solutions are difficult to obtain or require extensive calculations.[4] Both full-wave bridge and double-wye rectifiers were studied on the Anacom. Data taken on the Anacom substantiate the validity of the calculated curves and the mathematical expressions presented in this paper.

The full-wave bridge circuit was supplied from a 19-volt 3-phase source of low reactance, providing an essentially infinite bus. Anacom transformers are specially built on high-grade hypersil-steel cores, and require exciting currents which are negligible relative to the current base used in the laboratory study. These were connected to provide rectifier transformers with a line-to-line turns ratio of unity. The transformers have an inductance of 7 millihenrys per phase. Computer reactors of 93 millihenrys were connected in each phase, giving a total reactance of 37.7 ohms per phase.

One problem encountered in setting up the laboratory circuit was in finding rectifier elements having low voltage drop in the forward direction and very high impedance in the reverse direction. Selected selenium cells, about 2 inches in diameter, were found to have the best all-around characteristics. Forward and reverse volt-ampere curves for the cells used are shown in Fig. 6. The forward-direction volt-ampere curve is shown on an expanded scale in Fig. 7.

With a circuit which gave wave forms judged to be good when viewed on an oscilloscope, data shown on the regulation curve of Fig. 8 were taken. Data shown by crosses in Fig. 8 were corrected for voltage drop across the rectifier cells by adding the voltage drop corresponding to the load current, but no account was taken of losses in computer transformers and reactors. The total a-c resistance in the circuit used in the laboratory, including transformers, was 2.80 ohms per phase. Corrections were made for this resistance by dividing the watts loss in the a-c circuit by the current in the d-c circuit. In a 6-phase bridge, the current in the a-c circuit is 81.6 per cent of the load current. This is the conventional procedure of correcting for a-c circuit losses in rectifier circuits having low ratios of resistance to reactance. The circles in Fig. 8 were plotted from data corrected for voltage drop across the a-c circuit resistance. The solid line in Fig. 8 is the calculated voltage-regulation curve from Fig. 2.

Sample oscillograms made on the laboratory circuit are shown in Fig. 9. Oscillograms 9A, 9C, and 9E show the no-load line-to-line voltage on the rectifier side of the transformer, superimposed on the line-to-line voltage under load. Oscillograms 9B, 9D and 9F show anode-to-cathode voltages at no load and under load and a trace of the anode current under load.

Comparison of Laboratory and Theoretical Wave Shapes

To understand the operation of a full-wave bridge circuit more fully during the third mode of operation, it is helpful to deduce the voltage and current waves by physical reasoning. Line-to-line voltage, anode current, and anode-to-cathode voltage have been constructed for the case of 30 degrees inherent delay and 90 degrees overlap and the resulting wave forms compared with laboratory oscillograms.

Line-to-Line Voltage

Consider the circuit and vector diagrams shown in Fig. 3 and the corre-

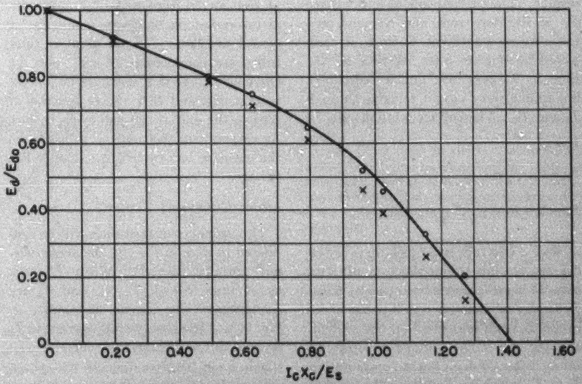


Fig. 8. **Voltage regulation curve for 6-phase rectifier showing laboratory data**

× Laboratory data

○ Laboratory data, corrected for resistance drop

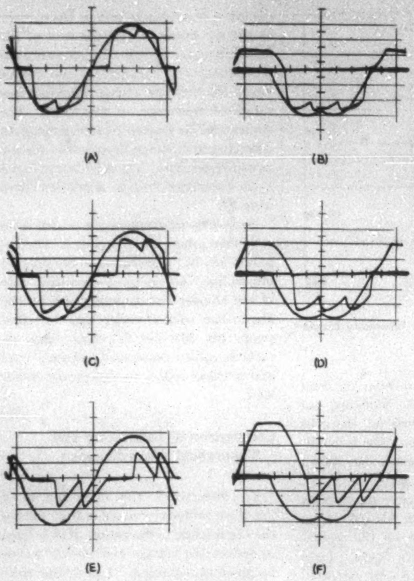


Fig. 9. Sample oscillograms from laboratory circuit

Left: line-to-line voltages

Right: anode-to-cathode voltages and anode currents

A and B: $\alpha=0$ degrees, $u=45$ degrees
C and D: $\alpha=0$ degrees, $u=60$ degrees
E and F: $\alpha=30$ degrees, $u=60$ degrees

sponding sine curves of Fig. 10A. For the condition of 30 degrees inherent delay, commutation from tube T_5 to tube T_1 will occur at an angle α, after no-load voltages on the two tubes are equal. This has been shown in Fig. 5. At the time $(t_1+\alpha)$, when this commutation begins, T_4 and T_6 are already conducting, and a temporary 3-phase short circuit results. The plotted line-to-line voltage E_{12} is therefore zero and remains zero until the commutation T_4 to T_6 is completed 30 degrees later, at time t_2. At time t_2, T_4 goes out, and a temporary line-to-line short circuit remains through T_1 and T_5. The voltage relations are

$$E_1=E_3$$

$$E_{13}=E_1-E_3=0$$

$$E_{12}=E_1-E_2=-\frac{3}{2}E_2$$

Since a new tube must pick up every 60 degrees, the line-to-line short circuit lasts only 30 degrees, until time $(t_2+\alpha)$ when T_2 begins to conduct. A 3-phase short circuit is again impressed on the system, and the line-to-line voltage collapses to zero. After 30 degrees the commutation T_5 to T_1, which has lasted for the overlap period of 90 degrees, is completed, and T_5 goes out. As T_1, T_2, and T_6 are still conducting, the voltage relations are

$$E_2=E_3$$

$$E_{23}=E_2-E_3=0$$

$$E_{12}=E_1-E_2=\frac{3}{2}E_1$$

As before, this line-to-line short circuit lasts 30 degrees until T_3 picks up at time $(t_3+\alpha)$ to impose another 3-phase short circuit. The line-to-line voltage remains at zero for 30 degrees until T_6 goes out, at the end of the 90-degree commutating period of T_6 to T_2. At time t_4 three tubes are conducting: T_1, T_3, and T_2. Obviously there is a short circuit on the voltage E_{12}, and the voltage remains at zero to the end of the half-cycle $(t_4+\alpha)$. An extension of this analysis will produce the negative half-cycle of the line-to-line voltage.

ANODE CURRENT

This sequence of operation will be considered to plot, in Fig. 10, the anode current through tube T_1. When T_1 picks up at time $(t_1+\alpha)$, T_4, T_5, and T_6 are conducting. The circuit is shown in Fig. 11A. The steady-state current in T_1 is equal to the 3-phase fault current $-I_c$, which is 90 degrees behind the phase voltage E_3 producing it and has a crest value of $(\sqrt{2}E)/X_c$. At time $(t_1+\alpha)$, therefore, the current $-I_c$ is at a negative maximum, as shown in Fig. 10B. Since the current i_1 cannot change instantly because of circuit inductance, the current wave $-I_c$ must be displaced to zero to properly describe i_1 in the interval $(t_1+\alpha)$ to t_2. At time t_2, tube T_4 cuts off and T_1, T_5, and T_6 are left conducting, as shown in Fig. 11B. The steady-state current in T_1 is now fault current I_a for a phase-a to phase-c fault. For such a fault, I_a lags its phase voltage E_1 by 120 degrees, or is in phase with E_2. It has a crest value of $(\sqrt{2}\sqrt{3}E_3)/2X_c$. Since circuit inductance forces i_1 just before t_2 to be equal to i_1 just after t_2, current I_a must be displaced upward to coincide with i_1 at time t_2 to properly describe the current in the interval t_2 to $(t_2+\alpha)$.

At $(t_2+\alpha)$, when T_2 picks up, T_1, T_5, and T_6 are already conducting, as shown in Fig. 11C. The current i_1 becomes 3-phase fault current I_a, which leads $-I_c$ by 60 degrees. As before, this current must be displaced, and that portion of the wave between $(t_2+\alpha)$ and t_3 is the current i_1. At time t_3, tube T_5 cuts off, leaving T_1, T_2, and T_6 conducting. The current is then equal to I_d for the next 30 degrees.

When tube T_3 picks up at $(t_3+\alpha)$, a 3-phase short circuit is again imposed through the paths illustrated in Fig. 11D. The current i_1 is 3-phase fault current I_a, and may be displaced and plotted for the interval $(t_3+\alpha)$ to t_4. Tube T_6 cuts off at t_4, leaving a phase-a to phase-b fault, Fig. 11E. Current i_1 is equal to current I_a for a phase-a to phase-b fault and leads 3-phase fault current I_a by 30 degrees, as plotted in Fig. 10B. With the proper displacement to make i_1 equal just before and just after t_4, I_a represents i_1 in the interval t_4 to $(t_4+\alpha)$. At time $(t_4+\alpha)$, tube T_4 begins conduction with T_1, T_2, and T_3 already conducting. The 3-phase fault is shown in Fig. 11F, and current i_1 is equal to $-I_b$, for a 3-phase fault. This current, leading 3-phase fault current I_a by 60 degrees, is plotted in Fig. 10B, and describes i_1 in the interval $(t_4+\alpha)$ to t_5, to complete the conduction period of 210 degrees.

ANODE-TO-CATHODE VOLTAGE

The anode-to-cathode voltage readily follows from the plot of line-to-line voltage. Fig. 10C shows the theoretical trace of anode-to-cathode voltage for tube T_1. Reference to the circuit of Fig. 3 shows that whenever T_3 is conducting, the anode-to-cathode voltage of T_1 will be the same as line-to-line voltage E_{13} plotted in Fig. 10A, neglecting, of course, tube drop. Tube T_3 conducts from time $(t_3+\alpha)$ to t_1, a total of 210 degrees. At t_1, tube T_3 goes out, leaving T_5, T_4, and

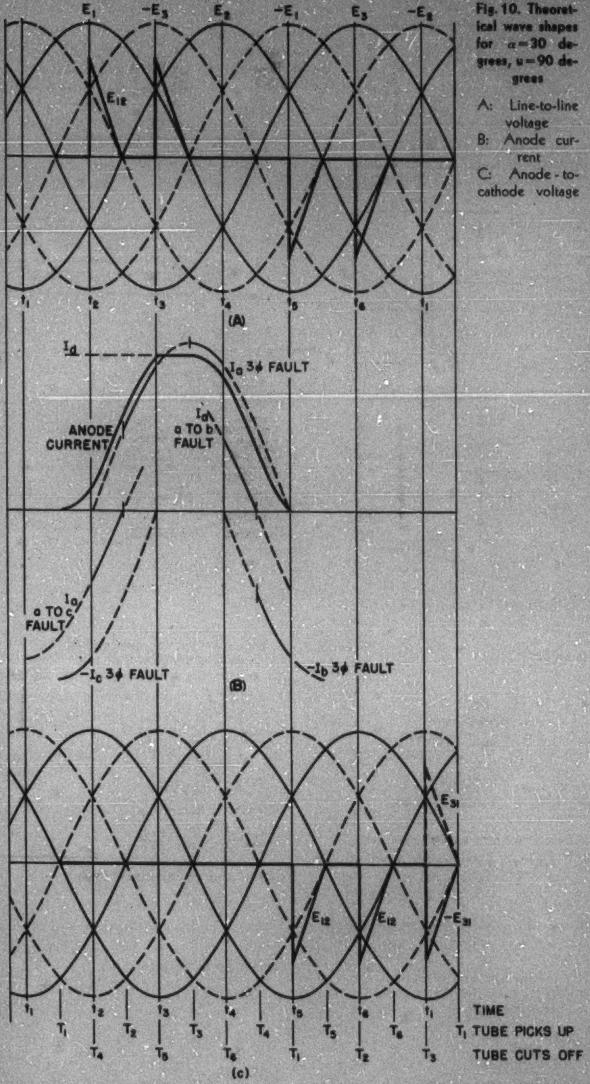


Fig. 10. Theoretical wave shapes for $\alpha=30$ degrees, $u=90$ degrees

A: Line-to-line voltage
B: Anode current
C: Anode-to-cathode voltage

T_6 conducting. The voltage across T_1 is then $E_1-E_3=E_{13}=-E_{31}$. From the consideration of a balanced 3-phase vector diagram, E_{31} is 120 degrees ahead of E_{13}. Since the voltage $-E_{31}$ appears across T_1 for 30 degrees, until T_1 picks up at $(t_1+\alpha)$, only 30 degrees of E_{31}, plotted as a dashed line, is shown. The negative of this voltage produces the third pip of the anode-to-cathode voltage in the interval t_1 to $(t_1+\alpha)$. After T_1 picks up at time $(t_1+\alpha)$, the anode-to-cathode voltage remains at zero to the end of the cycle at $(t_3+\alpha)$ or, more specifically, all the way to t_4, when T_1 stops conduction.

Laboratory Oscillograms

Oscillograms taken on the Anacom forceably illustrate the accuracy of the foregoing constructions. Fig. 12A shows the oscilloscope calibration, line-to-line voltage at no load, and a line-to-line voltage for 30 degrees inherent delay, with approximately 90 degrees overlap. The oscilloscope was adjusted to make one time division equal to approximately 36 degrees, and this calibration was maintained for all pictures. Fig. 12B illustrates a line-to-line voltage, such as E_{12} plotted in Fig. 10A, and a corresponding anode-to-cathode voltage such as that plotted in Fig. 10C. Fig. 12C shows a no-load anode-to-cathode voltage for reference, an anode-to-cathode voltage for a tube under load, and the current through the same tube.

Appendix. Delta 6-Phase Delta Double-Way Rectifier (Full-Wave Bridge)

This circuit is shown in Fig. 13. Resistance is assumed to be negligible, and transformer leakage reactance is shown external to an ideal transformer. The ratio of transformation between primary and secondary line-to-line voltages is chosen as unity. Line-to-neutral voltages behind source reactance are

$$E_a=\sqrt{2}E_s\cos\left(\omega t-\frac{\pi}{3}\right)$$

$$E_b=\sqrt{2}E_s\cos(\omega t-\pi)$$

$$E_c=\sqrt{2}E_s\cos\left(\omega t+\frac{\pi}{3}\right) \qquad (1)$$

where E_s is the rms value of transformer secondary no-load line-to-neutral voltage.

Line-to-neutral voltages applied to the anodes of one 3-phase rectifier group, and the cathodes of the other 3-phase rectifier group are E_1, E_2, and E_3. Normal commutation from tube T_3 to tube T_1 of rectifier group I occurs when their anode-to-cathode voltages are equal, that is, when $E_3=E_1$. Equations 1 have been selected with the time-axis reference such that commutation from T_3 to T_1 of rectifier group I normally starts when $t=0$.

Commutation in rectifier group II will occur when one cathode-to-anode voltage starts to become more negative than the others. This is the same as when the negative of the line-to-neutral voltage applied to one cathode becomes equal to the one on the conducting phase of group II. The negative of the line-to-neutral voltages applied to the cathodes of rectifier group II are $-E_1$, $-E_2$, and $-E_3$.

From the phasor diagram of Fig. 13, it can be seen that the two groups are displaced by 60 degrees. The circuit is equivalent to two 3-phase rectifiers in

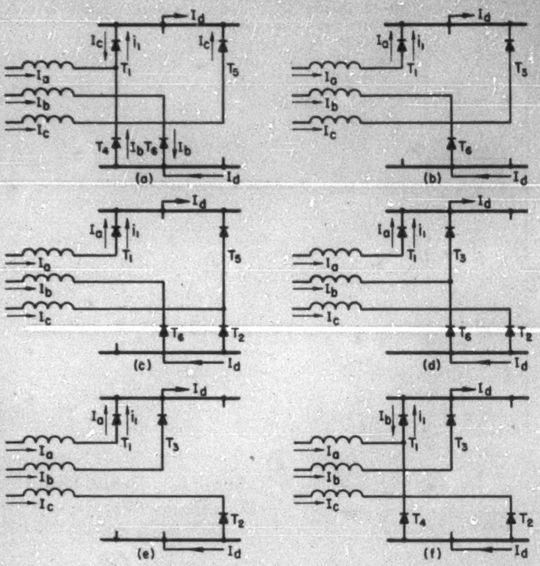


Fig. 11 (left). Circuit diagrams for the construction of the anode current in tube T_1

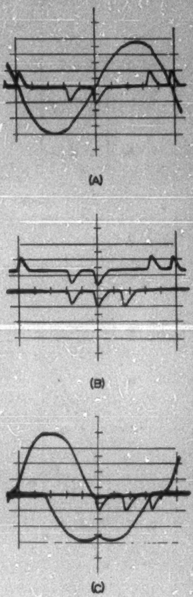


Fig. 12 (right). Laboratory wave shapes for $\alpha=30$ degrees, $u=90$ degrees

A: Line-to-line voltage
B: Line-to-line and anode-to-cathode voltages
C: Anode-to-cathode voltage and anode current

series with their alternating voltages displaced by 60 degrees. From the symmetry of the system, the average direct voltage can be obtained by figuring the average of the instantaneous voltage over a 60-degree period.

FIRST MODE OF OPERATION

Simultaneous conduction on two phases results in a temporary line-to-line short circuit on the secondary of the transformer. From Fig. 13

$$E_{ca}=Lpi_c-Lpi_a+E_{c'a'} \tag{2}$$

During a T_5 to T_1 commutation, $E_3=E_1$ and

$$E_{c'a'}=E_{31}=E_3-E_1=0 \tag{3}$$

Since T_6 is the only other tube conducting during this period, $i_c=-i_a$; therefore

$$E_{ca}=-2Lpi_a \tag{4}$$

During the first mode of operation, the instantaneous direct voltage has two distinct intervals, the commutation period and the noncommutation period. Measuring time from t_1 of Fig. 4, the instant of normal commutation from T_5 to T_1 on rectifier group I, the first interval will be from $\omega t=0$ at the start of commutation to $\omega t=u$ at the end of commutation, and the second interval will be from $\omega t=u$ to $\omega t=60$ degrees, when a commutation starts on rectifier group II. From Fig. 13, remembering that T_1, T_5, and T_6 are conducting, the instantaneous direct voltage is given by

$$e_d=-E_2+E_1=-E_{b'}+E_{a'}=E_{a'b'}$$

$$e_d=E_{ab}-Lpi_a+Lpi_b \tag{5}$$

But during the first interval, with T_6 being the only tube conducting in group II, $i_b=-I_d$ a constant; therefore from equation 4

$$e_{d1}=E_{ab}-Lpi_a=E_{ab}+\frac{E_{ca}}{2}$$

$$e_{d1}=\frac{3}{2}\sqrt{2}E_s\cos\omega t \tag{6}$$

During the second interval only T_6 and T_1 are conducting; $i_a=-i_b=I_d$, a constant. From equation 5

$$e_{d2}=E_{ab}=\sqrt{3}\sqrt{2}E_s\cos\left(\omega t-\frac{\pi}{6}\right) \tag{7}$$

The average direct voltage is

$$E_d=\frac{3}{\pi}\left[\int_0^u\frac{3}{2}\sqrt{\ }\ E_s\cos\omega t\,d(\omega t)+\int_u^{\pi/3}\sqrt{6}E_s\cos\left(\omega t-\frac{\pi}{6}\right)d(\omega t)\right] \tag{8}$$

Evaluating the two definite integrals between the indicated limits, collecting terms and reducing, there is obtained

$$E_d=\frac{3}{\pi}\sqrt{6}E_s\left(\frac{1+\cos u}{2}\right) \tag{9}$$

This is twice the value of a 3-phase single-way rectifier, whose average direct voltage is

$$E_d=\frac{3}{\pi}\sqrt{\frac{3}{2}}\,E_s\left(\frac{1+\cos u}{2}\right)$$

which verifies the statement that a single bridge circuit is equivalent to two 3-phase rectifiers in series. In addition, the 60-degree shift between the two groups of the bridge make it an equivalent 6-phase circuit.

The commutating angle u is obtained by integrating equation 4 and evaluating the integral at the boundaries $\omega t=0$, at which $i_a=0$, and $\omega t=u$, at which $i_a=I_d$

$$I_d=\int_0^{u/\omega}\frac{-E_{ca}}{2L}\,dt=\int_0^u\frac{\sqrt{3}\sqrt{2}}{2X_c}E_s\times\cos\left(\omega t-\frac{\pi}{2}\right)d(\omega t)$$

After integrating and rearranging

$$\cos u=1-\frac{2I_dX_c}{\sqrt{6}E_s} \tag{10}$$

Substituting equation 10 in 9

$$E_d=\frac{3}{\pi}\sqrt{6}E_s\left[1-\frac{I_dX_c}{\sqrt{6}E_s}\right] \tag{11}$$

The no-load direct voltage is obtained from equation 11 with $I_d=0$

$$E_{do}=\frac{3\sqrt{6}}{\pi}E_s \tag{12}$$

Dividing equation 11 by 12, the direct voltage in per unit of the no-load voltage is

$$\frac{E_d}{E_{do}}=1-\frac{I_dX_c}{\sqrt{6}E_s} \tag{13}$$

This is the conventional equation for the direct voltage ratio. The quantity I_dX_c/E_s is called the reactance factor.

SECOND MODE OF OPERATION

Following the T_5 to T_1 commutation in rectifier group I is a T_6 to T_2 commutation in rectifier group II. Commutation between

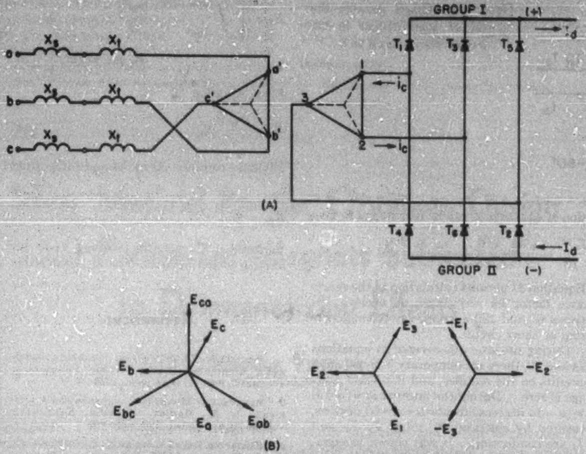


Fig. 13. Delta 6-phase delta double-way rectifier

A: Circuit diagram for bridge circuit B: Voltage phasor diagrams

these two starts when $-E_2=-E_3$ or when

$$-E_{b'c'}=-E_{23}=E_3-E_2=0 \qquad (14)$$

Normally this voltage becomes zero 60 degrees after the start of commutation from T_5 to T_1 on group I. From Fig. 13

$$E_{bc}=Lpi_b-Lpi_c+E_{b'c'} \qquad (15)$$

During commutation from T_5 to T_1, with T_6 being the only tube conducting in group II, from equation 4

$$E_{ca}=2Lpi_c \qquad (16)$$

Since $i_b=-I_d$, a constant, from equations 15 and 16, there results

$$-E_{b'c'}=-E_{bc}-\frac{E_{ca}}{2}$$

$$-E_{b'c'}=\frac{3\sqrt{2}}{2}E_s \cos \omega t \qquad (17)$$

If the commutation from T_5 to T_1 on group I tends to exceed 60 degrees, commutation from T_6 to T_2 on group II cannot start at its normal time since, according to equation (17) for $\omega t=60$ degrees, $-E_{b'c'}>0$. Thus, for reactance factors exceeding those for which $u=60$ degrees, commutation cannot start at normal time. Since commutation from T_6 to T_2 on group II cannot start at normal time, by the same reasoning, commutation from T_5 to T_1 on group I did not start at its normal time if there was still a commutation on group II. Thus, a forced delay in commutation is introduced for values of reactance factors greater than that which make $u=60$ degrees. From equation 10, this value is $\sqrt{6}/4$. When the reactance factor is greater than $\sqrt{6}/4$, the commutating angle u remains constant at 60 degrees, with forced delay in commutation being introduced. This can be verified by drawing sine waves with delayed commutation and an overlap angle u of least 60 degrees, remembering that symmetry must be retained.

The average direct voltage for this type of operation is obtained from equation 8, with the limits of integration being α and $\alpha+60$ degrees for the first integral; the second not being involved, since commutation occurs during the entire 60-degree period considered in obtaining the average direct voltage. The angle α is the angle of forced delay.

$$E_d=\frac{3}{\pi}\int_{\alpha}^{\alpha+\pi/3}\frac{3}{2}\sqrt{2}E_s \cos(\omega t)d(\omega t)$$

Evaluating this integral between the given limits and reducing terms

$$E_d=\frac{9\sqrt{2}}{2\pi}E_s \cos(\alpha+30°) \qquad (18)$$

Dividing equation 18 by 12

$$\frac{E_d}{E_{do}}=\frac{\sqrt{3}}{2}\cos(\alpha+30°) \qquad (19)$$

From equation 4

$$I_d=\int_{\alpha}^{\alpha+\pi/3}\frac{\sqrt{3}\sqrt{2}}{2X_c}E_s \cos\left(\omega t-\frac{\pi}{2}\right)d(\omega t)$$

After integrating and rearranging

$$\cos(\alpha+30°)=\sqrt{1-\left[\frac{2I_dX_c}{\sqrt{6}E_s}\right]^2} \qquad (20)$$

Rearranging equation 20

$$\frac{I_dX_c}{E_s}=\frac{\sqrt{6}}{2}\sin(\alpha+30°) \qquad (21)$$

Equation 21 gives reactance factors for angles of inherent delay between 0 and 30.

Third Mode of Operation

When $\alpha=30$ degrees the two voltages about to commutate become equal, even if there is still a commutation in the other rectifier group. For example, with a T_5 to T_1 commutation occurring in group I, the voltages in group II involved in the next commutation, $-E_3$ and $-E_2$, are equal at $\alpha=30$ degrees, as shown in Fig. 5. The commutation from T_6 to T_2 would therefore start with a delay of 30 degrees, even with reactance factors larger than $(3\sqrt{2})/4$. With increases in the reactance factor above $(3\sqrt{2})/4$, the delay angle is constant at 30 degrees, the overlap angle increases beyond 60 degrees, and both rectifier groups then commutate simultaneously for a part of the cycle.

Referring to Fig. 11A, when T_1 picks up at $(t_1+\alpha)$, tubes T_4, T_5, and T_6 are already conducting. Since T_2 is not conducting, $i_1=-i_c$, where i_1 is the instantaneous current in T_1. As there is a temporary 3-phase short circuit on the rectifier

$$i_a+i_b+i_c=0$$

and

$$E_{a'b'}=E_{b'c'}=E_{c'a'}=0$$

Therefore

$$E_{ca}=Lpi_c-Lpi_a$$

$$E_{ab}=Lpi_a-Lpi_b \qquad (22)$$

which give

$$2E_{ca}+E_{ab}=3Lpi_c=-3Lpi_1 \qquad (23)$$

or

$$pi_1=\frac{\sqrt{2}E_s}{L}\cos(\omega t-120°) \qquad (24)$$

Measuring time from t_1, equation 24 is valid for the interval $\omega t=\alpha=30$ degrees to $\omega t=(\alpha+\delta)=(u-30$ degrees), as shown in Fig. 14.

When $\omega t=(\alpha+\delta)$, T_4 cuts off. Then $i_1=i_a=-i_c$, as shown in Fig. 11B, since there is a temporary line-to-line short circuit on the rectifier. From Fig. 13, $E_{c'a'}=0$; therefore

$$E_{ca}=Lpi_c-Lpi_a=-2Lpi_a=-2Lpi_1 \qquad (25)$$

or

$$pi_1=\frac{-E_{ca}}{2L}=\frac{\sqrt{3}}{2}\left(\frac{\sqrt{2}E_s}{L}\right)\times \cos(\omega t-90°) \qquad (26)$$

Equation 26 is valid for the interval $\omega t=(\alpha+\delta)=(u-30$ degrees) to $\omega t=(t_2+\alpha)=90$ degrees.

When $\omega t=(t_2+\alpha)$, T_2 picks up. As shown in Fig. 11C, there is again a temporary 3-phase short circuit on the rectifier and $i_1=i_a$. Also

$$i_a+i_b+i_c=0$$

and

$$E_{a'b'}=E_{b'c'}=E_{c'a'}=0$$

Therefore

$$E_{ca}=Lpi_c-Lpi_a$$

$$E_{bc}=Lpi_a-Lpi_c \qquad (27)$$

which give

$$2E_{ca}+E_{bc}=-3Lpi_a=-3Lpi_1 \qquad (28)$$

or

$$pi_1=\frac{2E_{ca}+E_{bc}}{-3L}=\frac{\sqrt{2}E_s}{L}\cos(\omega t-60°) \qquad (29)$$

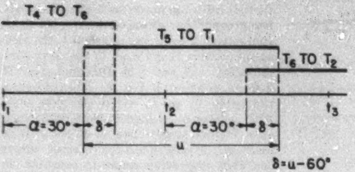


Figure 14. Graph showing the interval of commutations in the period from times t_1 to t_3

Equation 29 is valid for the interval $\omega t=(t_2+\alpha)=90$ degrees to $\omega t=(t_2+\alpha+\delta)=(u+30$ degrees).

When $\omega t=\alpha$, $i_1=0$ and when $\omega t=(t_2+\alpha+\delta)$, $i_1=I_d$. Using these boundaries to evaluate equations 24, 26, and 29

$$I_d=\left[\int_{\alpha}^{\alpha+\delta}\frac{\sqrt{2}\,E_s}{X_c}\cos(\omega t-120°)\times d(\omega t)+\int_{\alpha+\delta}^{t_2+\alpha}\frac{\sqrt{3}}{2}\left(\frac{\sqrt{2}\,E_s}{X_c}\right)\times \cos(\omega t-90°)d(\omega t)+\int_{t_2+\alpha}^{t_2+\alpha+\delta}\frac{\sqrt{2}\,E_s}{X_c}\cos(\omega t-60°)\,d(\omega t)\right] \quad (30)$$

From Fig. 14, $\alpha=30$ degrees, $\delta=u-60$ degrees, $t_2=60$ degrees; therefore $\alpha+\delta=u-30$ degrees, $t_2+\alpha=90$ degrees, and $t_2+\alpha+\delta=u+30$ degrees. Thus the solution of equation 30 can be rearranged to give

$$\frac{I_dX_c}{E_s}=\frac{\sqrt{2}}{2}+\frac{\sqrt{6}}{4}\sin u-\frac{\sqrt{2}}{4}\cos u \quad (31)$$

Equation 31 permits calculation of the reactance factor for any angle of overlap between 60 and 120 degrees, the latter occurring at short circuit.

During the intervals covered by equations 24 and 29, there are temporary 3-phase short circuits on the rectifier, and the direct voltage is zero. During the interval $\omega t=(\alpha+\delta)=(u-30$ degrees) to $\omega t=t_2+\alpha=90$ degrees, covered by equation 26, tubes T_1, T_5, and T_6 are conducting. It was shown in equation 6 that the direct voltage in this interval is

$$e_d=\frac{3\sqrt{2}}{2}E_s\cos\omega t \quad (32)$$

The average direct voltage is

$$E_d=\frac{3}{\pi}\left[\int_{u-30°}^{t_2+\alpha=90°}\frac{3\sqrt{2}}{2}\times E_s\cos\omega t\,d(\omega t)\right] \quad (33)$$

From equation 12

$$E_{do}=\frac{3\sqrt{6}}{\pi}E_s \quad (34)$$

Dividing equation 33 by 34 and integrating

$$\frac{E_d}{E_{do}}=\frac{\sqrt{3}}{2}[1-\sin(u-30°)] \quad (35)$$

Equation 35 gives the voltage ratio for overlap angles between 60 and 120 degrees.

References

1. Rectifier Transformer Characteristics, **AIEE Committee Report.** *AIEE Transactions*, volume 68, part II, 1949, pages 1305–11.

2. Equivalent Machine Constants for Rectifiers, **I. K. Dortort.** *Electrical Engineering*, volume 70, September 1951, page 802.

3. Rectifier Fault Currents, **C. C. Herskind, H. L. Kellogg.** *AIEE Transactions (Electrical Engineering)*, volume 64, 1945, pages 145–50, 441–42.

4. A Large-Scale General-Purpose Electric Analogue Computer, **E. L. Harder, G. D. McCann.** *AIEE Transactions*, volume 67, part I, 1948, pages 664–73.

◆

Discussion

I. K. Dortort (I-T-E Circuit Breaker Company, Philadelphia, Pa): May I offer my thanks to the authors of this excellent paper for the theoretical and laboratory corroboration they have provided for much of the material in my paper on extended regulation curves.[1] It should be pointed out that both papers were prepared independently and with no knowledge of the other.

The results presented in both papers apply only to truly free rectifier tubes, in which conduction can be initiated in any part of the cycle whenever the anode becomes more positive than the cathode. This precludes the use of impulse-fired igniters without holding electrodes or grids with an impulse shorter than 60 degrees.

The statement made in the summary of the paper that the equations derived are not affected by transformer connections is too broad. Without further proof the equations should be considered applicable only to the 6-phase double-way and double-wye rectifiers. I take it that the authors had only these two connections in mind.

It is also stated that reactance factors may be converted to per-unit reactance drop by multiplying by $2/\sqrt{6}$. This is true only for the first mode of operation and does not hold for modes 2 and 3. While this is hardly the time to bring the ASA Standard C 34.1[2] under discussion, this error confirms my earlier contention that the term "reactance factor" and its definition in ASA Standard C 34.1 was an unwise choice. To me the term connotes constancy, in the same way that the reactance of a transformer is constant. The term "load factor" used in my paper is not a happy choice, either. Its use does not imply constancy over a range of load, and does result in simpler expressions, but a better name is needed.

It is suggested that to facilitate practical computations equations 19 and 35 should be converted or expanded to include expressions in F_x (if we wish to stay with the reactance factor). Also equation 31 should be simplified to involve only a single trigometric term in u. These equations could be written as follows

$$\frac{E_d}{E_{do}}=\frac{1}{2}\sqrt{3}\sqrt{1-\left[\sqrt{\frac{2}{3}}F_x\right]^2} \quad (19)$$

$$F_x=\frac{1}{2}\sqrt{2}[1+\sin(u-30)] \quad (31)$$

$$\frac{E_d}{E_{d_o}}=\sqrt{3}\left[1-\frac{1}{2}\sqrt{2}F_x\right] \quad (35)$$

The oscillograms reproduced in Fig. 9 of the paper are of special interest in that they illustrate very clearly the existence of the delay angle α, without the positive-voltage loop that normally exists during this interval in phase-controlled rectifiers.

References

1. Extended Regulation Curves for 6-Phase Double-Way and Double-Wye Rectifiers, **I. K. Dortort.** *AIEE Transactions*, volume 72, part I, May 1953, pages 192–202.

2. Pool Cathode Mercury-Arc Power Converters. *ASA 34.1—1949*, American Standards Association, New York, N. Y., 1949.

R. L. Witzke, J. V. Kresser, J. K. Dillard: Several interesting observations have been made by Mr. Dortort, who has prepared a fine paper on this same subject.[1] His suggestions relative to equations 19, 31, and 35 are desirable for the reasons he has given. These equations also appear in the text of our paper, where equations describing the three modes of operation are summarized. The form suggested for equation 35 shows that the voltage ratio is a linear function of the reactance factor F_x (equal to I_cX_c/E_s in our notation) during the third mode of operation.

In our statement referring to transformer connections, we were thinking principally of connections used with 6-phase double-way and double-wye circuits. If other 6-phase connections are encountered, the specific circuit should be investigated separately.

The statement that reactance factors may be converted to per-unit reactance drop by multiplying by $2/\sqrt{6}$ has caused some misunderstanding. Our intention was to show the relationship between the reactance factors used by us, and the load factors used in an earlier paper by Mr. Dortort[2] and repeated in his current paper.[1] Multiplying the abscissas of our Fig. 2 by $2/\sqrt{6}$, we get the curve *abc*, given in Fig. 11 of the latter paper, all the way out to short circuit.

Neither of the papers presented here includes consideration of leakage reactance between secondary windings of double-wye connections. However, a detailed analysis of this interesting case is given in our discussion of Mr. Dortort's paper.[1] For most practical applications, this reactance is negligible.

It is appropriate at this time to mention some earlier work on regulation of rectifiers which came to our attention after our paper had been submitted for publication. A. J. Maslin published an excellent article[3] several years ago, in which he recognized the three modes of operation, and the same regulation characteristics given in our paper for the double-way circuit. The modes of operation for the double-wye circuit were also recognized by Prince and Vogdes.[4] We regret that we were not cognizant of these works during the preparation of our paper, for it would have been of inestimable value in setting up our laboratory study.

REFERENCES

1. See reference 1 of Mr. Dortort's discussion.
2. See reference 2 of the paper.
3. THREE-PHASE RECTIFIER CIRCUITS, A. J. Maslin. *Electronics*, New York, N. Y., December 1936, pages 28–31.
4. MERCURY-ARC RECTIFIERS AND THEIR CIRCUITS (book), Prince, Vogdes. McGraw-Hill Book Company, New York, N. Y., 1927.

New Method for the Optimum Design of Permanent Magnets Subjected to Demagnetizing Effects

HANS K. ZIEGLER
NONMEMBER AIEE

Synopsis: Extensive design investigations with the presently most important permanent-magnet materials have disclosed that optimum operating conditions for all practical interesting demagnetizing conditions are confined to minor hysteresis loops originating from a portion of the main demagnetization curve between the point of maximum-energy product and a point below, which differs for various materials. It has also been revealed that the results of design methods, based on the assumption of one constant slope for all minor loops, may be rather sensitive to the choice of the constant value. Combining both results in the use of an average slope value computed for the confined optimum range, new design relations for permanent magnets subjected to demagnetizing effects by varying either external reluctance or externally applied magnetomotive force have been developed. They can be utilized to derive analytically or graphically general design charts for each permanent-magnet material, permitting the direct reading of the optimum per-unit values of the magnet dimensions, depending upon the degree of demagnetization, and the selection of the material which offers the smallest size or the most suitable shape for a considered application. Interrelations for the two types of demagnetization have been developed.

THE development of improved permanent-magnet alloys, especially of the alnico type, has advanced the application of permanent magnets in many fields to an enormous extent. The demand for alnico-type magnets has reached proportions which cannot be satisfied by the limited allotment of necessary scarce materials, particularly cobalt and nickel, compatible with the overall national need for these materials. For the past several years, therefore, increased attention has been directed to the goal of utmost economical utilization of permanent-magnet alloys containing scarce materials, and to the development of new alloys which do not require critical components. The replacement of cut-and-try designs of magnets by systematic design methods to obtain optimum dimensions has been found indispensable for the conservation of materials, as well as for the evaluation of newly developed alloys. The problem of the optimum design of permanent magnets has been studied by several investigators.[1–5] This paper presents a method for the solution of the problem different from any used previously, which has been utilized since 1947 with very satisfactory practical results by the Signal Corps Engineering Laboratories. It lends itself particularly to a rather accurate analytical treatment, and therefore offers a tool for general design investigations which will be appreciated by many designers.

Elementary Assumptions

Concurrent with previous studies on the subject [1–5] the following assumptions will be used in this paper:

1. The magnet has uniform cross section and is uniform in properties and magnetic state.

2. The magnet has initially been magnetized to saturation, in the preferred axis if the material is anisotropic.

3. The minor hysteresis loop originating from the main demagnetization curve are of so small an area that they can be replaced for practical purposes by a straight line connecting their tips.

4. All leakage flux is assumed to come from the ends or the pole shoes of the magnets. Leakage is therefore regarded as a terminal factor and the optimum design is achieved for the required total magnetic flux Φ_t including leakage. Where excessive leakage originating from the lateral magnet faces invalidates such simplification, this part of the leakage is assumed to be replaced by an equivalent fictitious terminal leakage.

5. The main demagnetization curve of permanent-magnet materials can be closely approximated by hyperbolas, as derived by Watson[6] from Lamont's equation.[7] (Investigations indicate that this is true with very high accuracy for all alnico alloys.)

The form of the equation is chosen in accordance with Underhill[4]

$$\mathbf{B}=\frac{\mathbf{H}_c-\mathbf{H}}{\mathbf{H}_c-a\mathbf{H}}\mathbf{B}_r \tag{1}$$

whereby the materials constant a is defined as

$$a=2\sqrt{\frac{\mathbf{B}_r\mathbf{H}_c}{\mathbf{B}_o\mathbf{H}_o}}-\frac{\mathbf{B}_r\mathbf{H}_c}{\mathbf{B}_o\mathbf{H}_o} \tag{2}$$

where $\mathbf{B}_o$ and $\mathbf{H}_o$ = the magnetic induction and the magnetizing force corresponding to the point of maximum energy product
$\mathbf{B}_r$ = the residual induction
$\mathbf{H}_c$ = the coercive force
$\mathbf{B}$ and $\mathbf{H}$ = the magnetic induction and the magnetizing force for points on the main demagnetization curve

6. The centimeter-gram-second system of units is used throughout this paper (see nomenclature).

Nomenclature

a = materials constant, defined by equation 2
a_{op} = optimum magnet cross section per unit flux, centimeters^2 per maxwell
A = cross section of magnet, general, centimeters^2
A_m = cross section of magnet, optimum, centimeters^2
A_o = cross section of magnet, ideal, centimeters^2
b = materials constant, defined by equation 47
$\mathbf{B}$ = magnetic induction, general, gausses
$\mathbf{B}_o$ = magnetic induction, at point of maximum energy product, gausses

---

Paper 53-152, recommended by the AIEE Basic Sciences and Rotating Machinery Committees and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted October 17, 1952; made available for printing December 5, 1952.

HANS K. ZIEGLER is with the Signal Corps Engineering Laboratories, Fort Monmouth, N. J.

This paper represents part of the results of a research and development program covering the optimum design of permanent magnets, permanent-magnet generators, and related equipment conducted from 1947 to 1951 by the Signal Corps Engineering Laboratories.

The author wishes to express his appreciation to G. C. Uehrin for his assistance in calculating and preparing the diagrams of this paper.

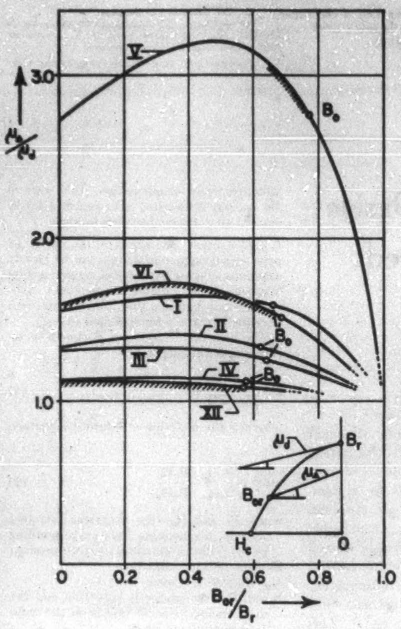


Figure 1 (left). Diagram showing variations of slope of minor loops for alnico alloys

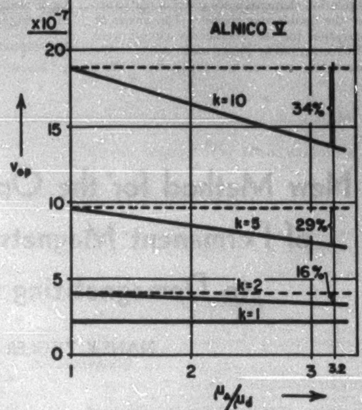


Figure 2 (right). Influence of choice of constant slope value upon optimum magnet volume for alnico V

$\mathbf{B}_{op}$ = magnetic induction, optimum, gausses
$\mathbf{B}_{or}$ = magnetic induction, at origin of optimum minor loop, gausses
$\mathbf{B}_r$ = magnetic induction, residual, gausses
c = demagnetization factor F_a/F_t
$\mathfrak{F}_a$ = magnetomotive force, externally applied, gilberts
$\mathfrak{F}_c$ = magnetomotive force, coercive of magnet M_s, gilberts
$\mathfrak{F}_{co}$ = magnetomotive force, coercive of magnet M_{so}, gilberts
$\mathfrak{F}_t$ = magnetomotive force, specified for normal operation, gilberts
$\mathfrak{F}_2$ = magnetomotive force, corresponding to point 2, Figures 5 and 7, gilberts
$\mathfrak{F}_3$ = magnetomotive force, corresponding to point 3, Figures 5 and 7, gilberts
$\mathbf{H}$ = magnetizing force, general, oersteds
$\mathbf{H}_c$ = magnetizing force, coercive, oersteds
$\mathbf{H}_o$ = magnetizing force, at point of maximum energy product, oersteds
$\mathbf{H}_{op}$ = magnetizing force, optimum, oersteds
k = demagnetization factor $\mathcal{R}_o/\mathcal{R}_t$
l = length of magnet, general, centimeters
l_m = length of magnet, optimum, centimeters
l_o = length of magnet, ideal, centimeters
l_{op} = length of magnet, optimum per unit, centimeters per gilbert
m_s = designation of minor loop corresponding to shape factor s, Figures 5 and 7
M_s = designation of main demagnetization curve corresponding to shape factor s and Φ_r, Figures 5 and 7
M_{so} = designation of main demagnetization curve corresponding to shape factor s and $\Phi_{ro} = \Phi_t$, Figures 5 and 7
$\mathcal{R}_o$ = open-circuit reluctance, gilberts per maxwell
$\mathcal{R}_t$ = operating-circuit reluctance, gilberts per maxwell
$\mathcal{R}_{ta}$ = designation for parallel displaced characteristic $\mathcal{R}_t$, Figures 5 and 7
$\mathcal{R}_a$ = designation of shifting auxiliary line, Figures 5 and 7
s = shape factor of magnet, general, per centimeter
s_m = shape factor of magnet, optimum, per centimeter
s_o = shape factor of magnet, ideal, per centimeter
v = magnet volume per unit external magnetic energy, general, centimeters^3 per maxwell-gilbert
v_{op} = magnet volume per unit external magnetic energy, optimum, centimeters^3 per maxwell-gilbert
V = magnet volume, general, centimeters^3
V_m = magnet volume, optimum, centimeters^3
V_o = magnet volume, ideal, centimeters^3
z = terms defined by equations 23 and 24
z_m = terms defined by equations 23 and 24
η = volume correction factor, general
η_m = volume correction factor, optimum
λ = magnification factor (ratio of magnetic inductions), general
λ_m = magnification factor (ratio of magnetic inductions), optimum
μ_a = average slope of minor loops for defined zone and shape factor 1, gausses per oersted
μ_{as} = resultant average slope of minor loops for defined zone and shape factor s, maxwells per gilbert
μ_d = differential permeability at the point $\mathbf{B}_r$, gausses per oersted
μ_Δ = slope of straight line replacement for minor loop, gausses per oersted
Φ_r = residual flux of magnet M_s, maxwells
Φ_{ro} = residual flux of magnet M_{so}, maxwells
$\tan^{-1} \mu_{as}$ = angle of inclination of minor loop corresponding to slope μ_{as}
I, II, III, IV, V, VI, XII = alnico grades

Considerations Regarding Slope of Minor Loops

Depending upon the type of magnetic alloy, the slope of the minor loops may vary considerably with the point of their origin at the main demagnetization curve. The actual variations can be considered in graphical or analytical point-by-point methods of solutions.[1,5] Complete ana-

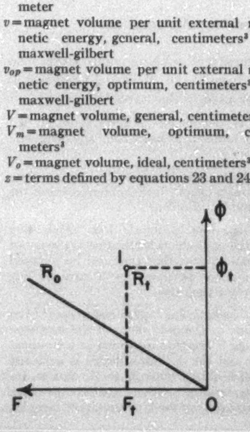


Figure 3 (left). Problem specification for demagnetization by varying external reluctance

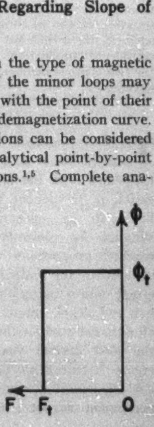


Figure 4 (right). Main demagnetization curve of hypothetical ideal magnet lytical methods require the knowledge of analytical relations for the slope variations, or the assumption of one constant slope, which must be chosen to approach as closely as possible the results obtained under consideration of the actual variations.

The conditions for the various alnico grades are qualitatively illustrated in Figure 1, which is based on manufacturers' quotations of slope values. The slope of the straight-line replacement for the minor loops originating at the main demagnetization curve and ending at the **B**-axis is designated with μ_Δ. The slope of the main demagnetization curve at the point of residual induction $\mathbf{B}_r$ is denoted with μ_d. The diagram exhibits the changes of the ratio μ_Δ/μ_d versus the ratio $\mathbf{B}_{or}/\mathbf{B}_r$ pertinent to the magnetic induction $\mathbf{B}_{or}$ at the point of origin of the minor loops. With the exception of the grades of alnico IV and XII, the variations are quite large and do not suggest the assumption of a constant slope. The previously proposed adoption[4] of a constant slope μ_Δ equal to the differential permeability μ_d at the point $\mathbf{B}_r$, that is $\mu_\Delta/\mu_d=1$, does not appear justified for each type of alloy, unless it is proved that the influence of the variations upon optimum design dimensions is within the accuracy limitations which must be allowed for any permanent-magnet design method because of the many other simplifying assumptions and variations in the materials charactersitics.

The influence of the choice of the value of a constant slope μ_Δ upon the optimum size of the magnet for alnico V material is illustrated in Figure 2. The diagram exhibits the optimum volume v_{op} of magnetic material per unit of externally available magnetic energy $\Phi_t \mathfrak{F}_t$ for various degrees $k=\mathcal{R}_o/\mathcal{R}_t$ of demagnetization by reluctance increases from $\mathcal{R}_t$ to $\mathcal{R}_o$, depending upon the choice of the value μ_Δ/μ_d, which according to Figure 1 can be taken for alnico V between 1 and 3.2. A practically linear decrease of the magnet volume with the increase of μ_Δ/μ_d is indicated, increasing in magnitude with increasing ratio of k. Maximum volume variations of 16 per cent to 34 per cent, for practical k values from 2 to 10, are affected between the border values 1 and 3.2 of μ_Δ/μ_d. These volume variations are too excessive to be negligible even for design methods claiming only limited accuracy. Design methods based on the assumption of one constant slope when applied to magnetic materials with considerable variations of the slope, therefore require particular considerations in the selection of a proper slope value to obtain optimum design accuracy.

As a first approach towards this goal, the arithmetical average of the slope of all minor loops originating from the main demagnetization curve could be introduced. Investigations started from this assumption have, however, uncovered general physical relations which permit a closer approach.

For all the permanent-magnet materials investigated as of the present, comprising all alnico grades and a number of newer materials, it was found that all optimum designs for any chosen degree of demagnetization are materialized for operating conditions which are associated with minor loops originating from a limited zone of the main demagnetization curve. The upper border of this limited zone was found for all materials to be equal to the point of maximum-energy product, corresponding to a magnetic induction $\mathbf{B}_o$, while the lower border varies with the type of material and the range of considered degrees of demagnetization, and may be located between $\mathbf{B}_{or}<\mathbf{B}_o$ and $\mathbf{B}_{or}=0$. The discovery of the insignificance of minor loops originating from points above the maximum-energy product for optimum designs is of great importance. Previous investigations[5] have revealed that it appears a general rule for all permanent-magnet alloys that the slope of the minor loops is subjected to its major changes for magnetic inductions above $\mathbf{B}_o$. This rule can also be seen proved by the steep descent of the functions between the points marked $\mathbf{B}_o$, which correspond to $\mathbf{B}_{or}/\mathbf{B}_r=\mathbf{B}_o/\mathbf{B}_r$ and $\mathbf{B}_{or}/\mathbf{B}_r=1$ in Figure 1. Permitting the designer to disregard these portions then leaves him with the remaining flatter portion of the functions between $\mathbf{B}_{or}/\mathbf{B}_r=\mathbf{B}_o/\mathbf{B}_r$ and $\mathbf{B}_{or}/\mathbf{B}_r=0$, and also in a much better position to derive an average value which does not deviate much from the maximum and minimum values. This condition is further improved by the observation, that is obviously a regularity, that magnetic materials showing large variations of the slopes, such as alnico V alloy, have their lower border closer to $\mathbf{B}_o$ than magnetic materials with smaller slope variations, which in general have their lower border closer to, or at, $\mathbf{B}_{or}=0$. The range of slope values to be considered for materials of large variations is thus limited to a still smaller zone.

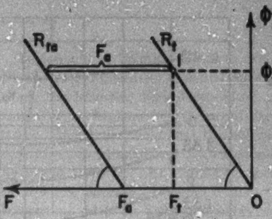


Figure 6. Problem specification for demagnetization by external applied magnetomotive force

An illustration of the results explained in the preceding is given in the design charts, Figures 12 and 17, which contain, for alnico V, VI, and XII, the relations for the magnetic induction $\mathbf{B}_{or}$ at the origin of those minor hysteresis loops which lead to optimum designs, depending upon

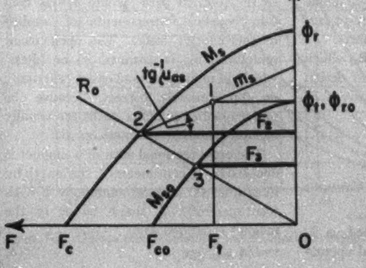
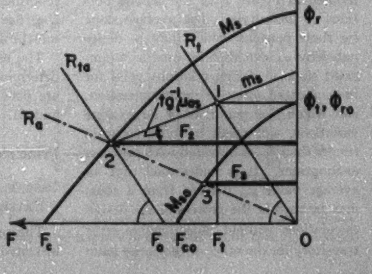


Figure 5 (left). Diagram showing basic geometrical relations for magnetization by varying external reluctance

Figure 7 (right). Diagram showing basic geometrical relations for demagnetization by external applied magnetomotive force



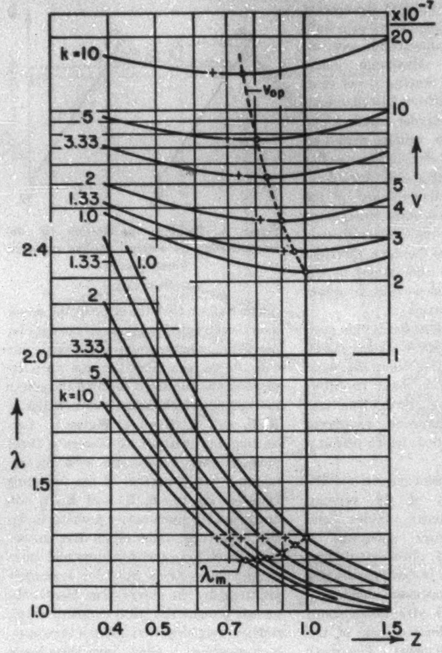


Figure 8. Design chart, containing v and λ functions versus z for alnico V, case A

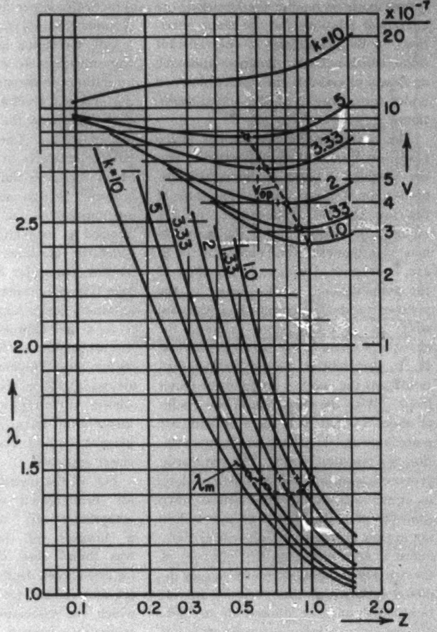


Figure 9. Design chart, containing v and λ functions versus z for alnico VI, case A

the degree of demagnetization k and c respectively. In no case does $\mathbf{B}_{or}$ exceed $\mathbf{B}_o$. For alnico V, a definite zone is indicated, approaching a maximum range extending between $\mathbf{B}_{or}=\mathbf{B}_o$ and $\mathbf{B}_{or}\approx 0.85\mathbf{B}_o$. For alnico VI and XII, larger zones are indicated, increasing with the degree of demagnetization. For practical purposes only degrees of demagnetization k and c below 10 are generally of interest, which for alnico VI and XII results in relatively small zones. For alnico V, VI, and XII, the zones for $1 \leq k \leq 10$ and $0 \leq c \leq 10$ are marked in Figure 1 by the shaded portions of the curves. The average slope for these portions, which may be designated by μ_a, can then be used as a constant slope in design relations valid for the indicated range of demagnetization, and which, for materials of large slope variations, will result in more accurate results than those obtained with previous assumptions. It is understood that an average value μ_a, derived from a zone, may be used in the process of the design calculations without objectionable error transitory for the range of the entire main demagnetization curve, if the evaluation of results is finally restricted to the zones to which it actually applies. For an initial investigation of a newly developed permanent-magnet alloy, the zone can be established by first using an average value μ_a computed for the range between $\mathbf{B}_{or}=\mathbf{B}_o$ and $\mathbf{B}_{or}=0$, and then to further limit the range in accordance with the results of these design calculations. The calculations should then be repeated with the improved average value computed from the zone. Hardly more than one repetition will ever be necessary to define the zone with sufficient accuracy to permit the development of design charts.

The relations for the values $\mathbf{B}_{or}$ which determine the zones of optimum design are listed in Appendix II, equations 50 and 51.

Principle of Design Method

Case A: Demagnetization by Varying External Reluctance

The problem should be defined as illustrated in Figure 3. A normal operating point, 1, calling for a total magnetic flux Φ_t and a total magnetomotive force $\mathfrak{F}_t$, to which a total magnetic reluctance $\mathfrak{R}_t$ corresponds, and a demagnetizing condition caused by an open-circuit magnetic reluctance $\mathfrak{R}_o$, represent the problem specifications. No restrictions are imposed upon the operating reluctance $\mathfrak{R}_t$ regarding its constancy for values of magnetic flux other than specified, which means the normal external magnetic circuit may be linear or nonlinear. R_t may be considered also an apparent reluctance, given by the quotient $\mathfrak{F}_t/\Phi_t$, where $\mathfrak{F}_t$ may contain constant components of applied magnetomotive force. The open-circuit reluctance $\mathfrak{R}_o$ is assumed to be linear, which, in many practical cases, is actually or approximately correct, because the major open-circuit flux paths are usually located in nonferromagnetic media.

The cross-sectional area of a magnet is denoted by A(centimeters^2), the length by l(centimeters) and the volume by $V=Al$-(centimeters^3). A shape factor of the magnet is defined as

$$s=l/A \tag{3}$$

For the optimum magnet, the corresponding values are designated by A_m, l_m, s_m, and V_m.

The degree of demagnetization is characterized by

$$k=\mathcal{R}_o/\mathcal{R}_t \qquad (4)$$

The standard unit volume of a magnetic material, $V=1$ centimeter^3, $A=1$ centimeter^2, and $l=1$ centimeter, has a shape factor $s=1$. The average slope of the minor loops, as defined before, for this standard unit is denoted by μ_a. As for the assumed uniformity of conditions, the magnetic characteristics of a magnet of any other size can be derived from those of the standard unit by variations in proportions to the dimensions, so it can be concluded that the resultant slope μ_{as} depends exclusively upon the shape factor s of the magnet

$$\mu_{as}=\mu_a/s \qquad (5)$$

Previous design methods[1-3,5] were based on the derivation of magnetic energy output curves for various constant degrees of demagnetization, whereby the maximum output was the indication for the optimum operating conditions. The method presented here is based on the introduction of a hypothetical ideal magnet the dimensions of which are subsequently corrected for by factors accounting for the deviations in characteristics between the ideal and the actual magnetic material. The minimum of the volume correction factor signifies the optimum design. For the hypothetical ideal magnet, it is assumed that the magnetic material under consideration has the same values $\mathbf{B}_r$ and $\mathbf{H}_c$ as the actual material, but that the main demagnetization curve is a perfect rectangle as shown in Figure 4, and that no minor hysteresis loops exist. For such an ideal magnet, which would not require a consideration of the demagnetization caused by the open-circuit reluctance $\mathcal{R}_o$, the dimensions would be simply given as

$$A_o=\Phi_t/\mathbf{B}_r \qquad (6)$$

$$l_o=\mathfrak{F}_t/\mathbf{H}_c \qquad (7)$$

$$V_o=\Phi_t\mathfrak{F}_t/(\mathbf{B}_r\mathbf{H}_c) \qquad (8)$$

The shape factor s_o of the ideal magnet is then

$$s_o=l_o/A_o \qquad (9)$$

To derive the correction factors to modify the ideal dimensions to actual dimensions, the following logical procedure is employed, as explained by means of Figure 5. It is understood that the normal operating point 1 must be located at minor hysteresis loops m_s. An infinite number of different slopes μ_{as} between 0 and ∞, concurrent with an infinite number of different sizes of magnets, which all satisfy the specifications, may be chosen, but according to equation 5, each different slope corresponds for a given material to a definite shape factor s of a magnet. For any chosen value of μ_{as}, a main demagnetization curve M_{so} of the given material for the corresponding shape factor s can be placed into the diagram, for instance, with a scale for which the residual flux Φ_{ro} is equal to Φ_t. A magnet with a main demagnetization curve M_{oo} does not satisfy the requirements, as M_{so} does not pass through intersection 2 between the minor loop m_s and the open-circuit line $\mathcal{R}_o$, but the curve M_{so} must only be magnified by a factor λ which is given by the ratio of

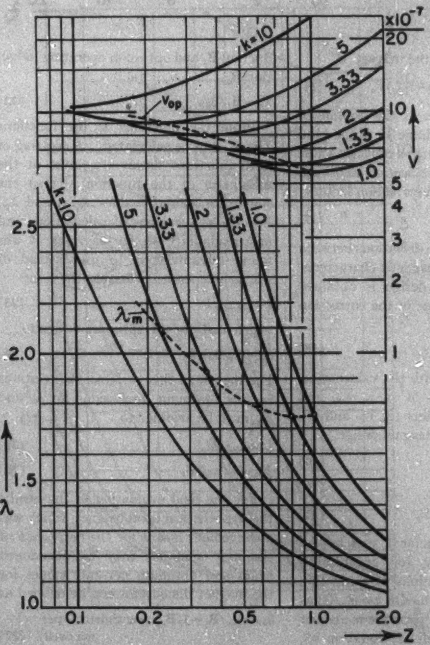


Figure 10. Design chart, containing v and λ functions versus z for alnico XII, case A

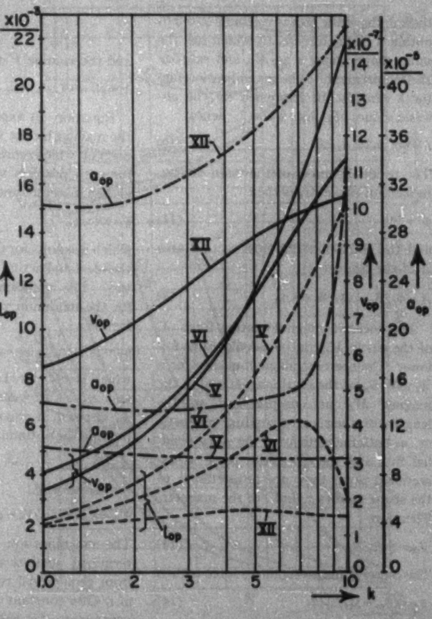


Figure 11. Design chart, containing optimum per-unit magnet dimensions, depending upon degree k of demagnetization, case A, for alnico V, VI, and XII

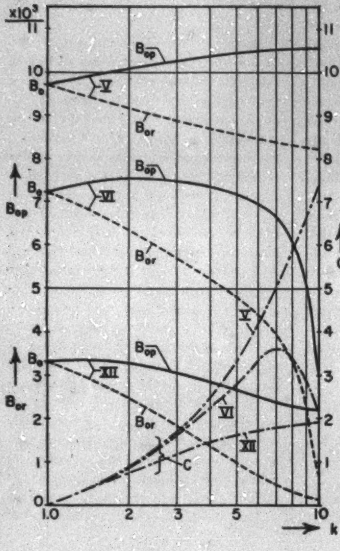


Figure 12 (left). Design chart, containing optimum values B_{op} and B_{or} depending upon degree of demagnetization k, case A, and equivalent values c, for alnico V, VI, and XII

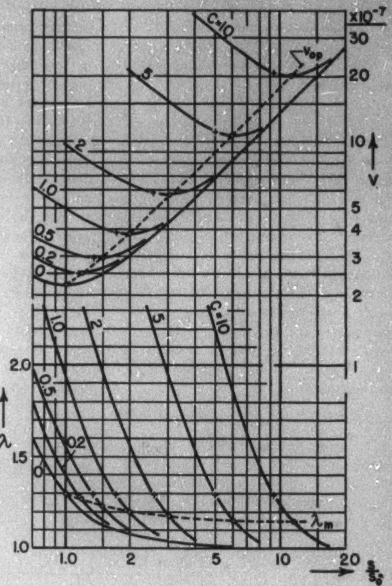


Figure 13 (right). Design chart, containing v and λ functions versus s/s_0 for alnico V, case B

the lengths of the radius vectors 0-2/0-3 to obtain a characteristic M_s which has the same shape factor s as M_{s0} and satisfies the requirement. The magnification factor λ can also be expressed by the abscissa values of points 2 and 3, hence

$$\lambda = 0\text{-}2/0\text{-}3 = \mathfrak{F}_2/\mathfrak{F}_3 \tag{10}$$

The residual flux Φ_r of the main demagnetization curve M_s is then

$$\Phi_r = \lambda \Phi_t \tag{11}$$

and the corresponding cross-section area A of the magnet is

$$A = \Phi_r/\mathbf{B}_r = \lambda \Phi_t/\mathbf{B}_r = \lambda A_o \tag{12}$$

The coercive magnetomotive force $\mathfrak{F}_c$ of the curve M_s can be calculated as follows: The coercive magnetomotive force $\mathfrak{F}_t$ and $\mathfrak{F}_{c0}$ of the ideal magnet and of the magnet M_{s0} are proportional to the lengths of their corresponding magnets, or, as both magnets have the same residual flux and therefore the same cross-sectional area, they are proportional to the shape factors s_o and s of the magnets. Therefore

$$\mathfrak{F}_{c0} = (s/s_o)\mathfrak{F}_t \tag{13}$$

The coercive magnetomotive force is then

$$\mathfrak{F}_c = \lambda \mathfrak{F}_{c0} = \lambda (s/s_o)\mathfrak{F}_t \tag{14}$$

The corresponding length l of the magnet is therefore

$$l = \mathfrak{F}_c/\mathbf{H}_c = \lambda (s/s_o)(\mathfrak{F}_t/\mathbf{H}_c) = \lambda (s/s_o) l_o \tag{15}$$

and the volume V of the magnet is

$$V = Al = \lambda^2 (s/s_o) A_o l_o = \lambda^2 (s/s_o) V_o \tag{16}$$

Equation 16 expresses the volume of the magnet by the volume V_o of the ideal magnet, which represents a constant value for each problem specification and each chosen material, and a correction factor

$$\eta = \lambda^2 (s/s_o) \tag{17}$$

which accounts for the deviations between the ideal and actual magnetic characteristics. The optimum design is obtained for the minimum value of the correction factor

$$\eta_m = [\lambda^2 (s/s_o)]_{\min} = \lambda_m^2 (s_m/s_o) \tag{18}$$

which is associated with the values s_m/s_o and λ_m. The values of s_m/s_o, λ_m, and η_m inserted in equations 12, 15, and 16, provide the optimum magnet design

$$A_m = \lambda_m A_o \tag{19}$$

$$l_m = \lambda_m (s_m/s_o) l_o \tag{20}$$

$$V_m = \lambda_m^2 (s_m/s_o) V_o = \eta_m V_o \tag{21}$$

The magnification factor λ can be determined analytically (or graphically) from geometrical relations as a function of s/s_o for constant values of $k = \mathcal{R}_o/\mathcal{R}_t$ as parameters, for each permanent-magnet material; see Appendix I, equation 48. As readily seen from equation 12, λ_m is physically the ratio between residual induction $\mathbf{B}_r$ and optimum operating induction $\mathbf{B}_{op}$.

$$\lambda_m = \mathbf{B}_r/\mathbf{B}_{op} \tag{22}$$

From the functions λ, the functions $\lambda^2(s/s_o)$ can be calculated. A plotting of both functions versus s/s_o gives at the minumum of the function $\lambda^2(s/s_o)$ the sought-for pair of values s_m/s_o and λ_m. For the convenience of calculations and best readability of the diagrams, it was found advantageous to use instead of s/s_o as independent variable, the term

$$z = (s/s_o)/k \tag{23}$$

which leads for the optimum design to

$$z_m = (s_m/s_o)/k \tag{24}$$

For the use of this variable, the relations for the optimum dimensions for l_m and V_m must be modified to

$$l_m = \lambda_m z_m k l_o \tag{25}$$

$$V_m = \lambda_m^2 z_m k V_o \tag{26}$$

For the final evaluation of the results the hypothetical ideal magnet, which was a convenient means for the derivation of the new relations, may be abandoned again and optimum per-unit values for the magnet dimensions can be defined as

$$a_{op} = \lambda_m/\mathbf{B}_r = 1/\mathbf{B}_{op} \text{ (centimeters}^2\text{ per maxwell)} \tag{27}$$

$$l_{op} = \lambda_m z_m k/\mathbf{H}_c = 1/\mathbf{H}_{op} \text{ (centimeters per gilbert)} \tag{28}$$

$$v_{op}=\lambda_m{}^2 z_m k/(\mathbf{B}_r\mathbf{H}_c)=1/(\mathbf{B}_{op}\mathbf{H}_{op})\text{ (centimeters}^3\text{ per maxwell-gilbert)} \qquad (29)$$

With these definitions, the optimum magnet dimensions are then

$$A_m=\Phi_t a_{op} \qquad (30)$$

$$l_m=\mathfrak{F}_t l_{op} \qquad (31)$$

$$V_m=\Phi_t\,\mathfrak{F}_t v_{op} \qquad (32)$$

In this form of representation the results of the design method are most convenient for the designer.

CASE B: DEMAGNETIZATION BY EXTERNALLY APPLIED MAGNETOMOTIVE FORCE

This problem should be defined as illustrated in Figure 6. A normal operating point 1 (Φ_t, $\mathfrak{F}_t$; $\mathfrak{R}_t$) should be specified as in case A. Demagnetization is affected by an externally applied magnetomotive force $\mathfrak{F}_a$, and the degree of demagnetization is defined by the ratio

$$c=\mathfrak{F}_a/\mathfrak{F}_t \qquad (33)$$

Deviating from case A, it is assumed that the normal operation circuit reluctance $\mathfrak{R}_t$ is constant and that the circuit under normal operation, point 1, does not contain an externally applied magnetomotive force. Accordingly the characteristic of $\mathfrak{R}_t$ is represented by a straight line through origin O and point 1, which under demagnetizing condition is displaced to the left by a distance corresponding to $\mathfrak{F}_a$, and parallel to its original position as signified by the line $\mathfrak{R}_{ta}$.

The derivation of the design relations follows exactly the same line of thought as explained in case A. For the derivation of the magnification factor λ, it is necessary to add a shifting auxiliary line $\mathfrak{R}_a$, connecting point 2 with the origin O, as demonstrated in Figure 7. The function λ can again be obtained analytically, or graphically, as a function of s/s_o, for constant values of parameters c for each permanent-magnet material; see Appendix I, equation 49.

The introduction of a variable corresponding to z, equation 23, was found unnecessary. Equations 19, 20, and 21 for the optimum dimensions of case A can therefore be utilized without modification. The optimum per-unit dimensions, except for a_{op}, equation 27, change however to

$$l_{op}=\lambda_m(s_m/s_o)/\mathbf{H}_c \qquad (34)$$

$$v_{op}=\lambda_m{}^2(s_m/s_o)/(\mathbf{B}_r\mathbf{H}_c) \qquad (35)$$

The final equations 20, 31, and 32 remain the same for cases A and B.

Demonstrative Examples of Design Charts

The results obtained with the design method presented are illustrated in design charts for alnico V ($\mathbf{B}_r=12{,}500$ gausses, $\mathbf{H}_c=600$ oersteds, $\mathbf{B}_o=9{,}680$ gausses, $\mathbf{H}_o=465$ oersteds, $a=0.916$, $\mu_a=-5.0$ gausses per oersted), alnico VI ($\mathbf{B}_r=10{,}500$ gausses, $\mathbf{H}_c=750$ oersteds, $\mathbf{B}_o=7{,}200$ gausses, $\mathbf{H}_o=515$ oersteds, $a=0.791$, $\mu_a=-5.0$ gausses per oersted) and alnico XII ($\mathbf{B}_r=5{,}800$ gausses, $\mathbf{H}_c=950$ oersteds, $\mathbf{B}_o=3{,}300$ gausses, $\mathbf{H}_o=538$ oersteds, $a=0.427$, $\mu_a=-3.7$ gausses per oersted).

For demagnetization by varying external reluctance, Figures 8, 9, and 10 represent the values v and λ as functions of the variable z for demagnetization factors k between 1 and 10. The value v is the volume of magnet per-unit external mag-

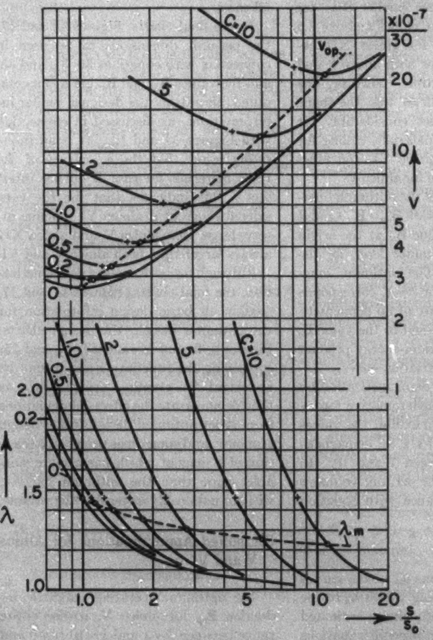


Figure 14. Design chart, containing v and λ functions versus s/s₀ for alnico VI, case B

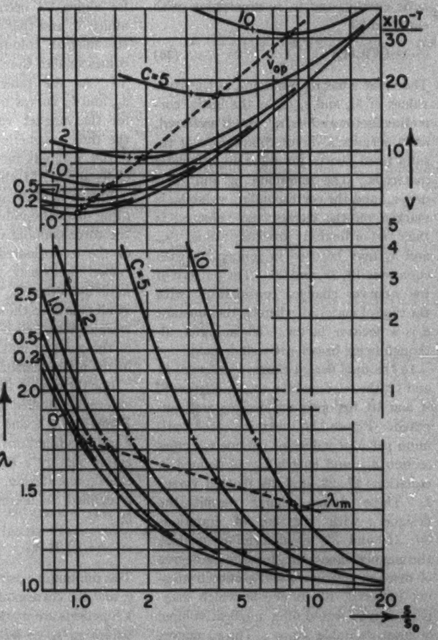


Figure 15. Design chart, containing v and λ functions versus s/s₀ for alnico XII, case B

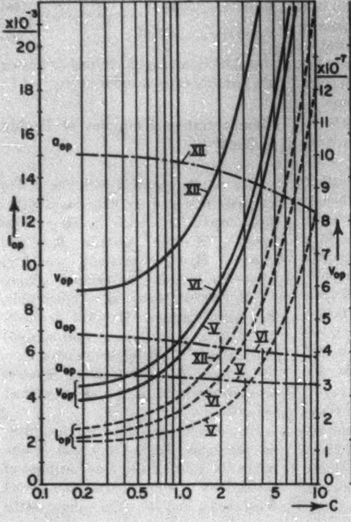
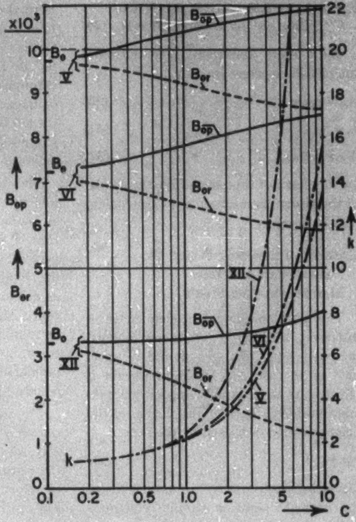


Figure 16 (left). Design chart, containing optimum per unit magnet dimensions depending upon degree of demagnetization c, case B, for alnico V, VI, and XII

Figure 17 (right). Design chart, containing optimum values B_{op} and B_{or} depending upon degree of demagnetization c, case B, and equivalent values k, for alnico V, VI, and XII



netic energy, and is, in accordance with equation 29

$$v=\lambda^2 zk/(B_rH_c) \qquad (36)$$

The value v has its minimum at the same values of λ_m and s_m/s_o as the initial correction factor $\eta=\lambda^2(s/s_o)$, and is preferred, as it permits a direct comparison of results from charts for different magnetic materials. The minimum per-unit volume v_{op} and the corresponding value λ_m is marked on the curves, the value z_m is their co-ordinated abscissa value. λ_m and z_m may be used in connection with equations 19, 25, and 26. The pattern of the v-curves changes considerably with the type of material, whereby the constant a is a decisive factor. The minima, although being broad, are well defined.

In the final design charts of Figures 11 and 12, the results of the charts, Figures 8, 9, and 10, for optimum designs are computed. Figure 11 contains the optimum per-unit values of volume v_{op}, cross section a_{op}, and length l_{op}, as defined in equations 27, 28, and 29, plotted versus k. These design charts permit the designer, with the use of equations 30, 31, and 32, to determine at once the magnet dimensions for a given degree of demagnetization k and a chosen magnet material, or to decide which magnet material would offer smallest volume or most desirable shape. The v_{op} curves of the charts reveal, for instance, that for values of k above 4.6 the magnet volume for alnico VI becomes smaller than for alnico V, and that for values k above 7.5 the magnet volume for alnico XII becomes smaller than for both alnico V and VI. At the same time, however, the a_{op} and l_{op} curves indicate that the shape for the magnet varies considerably for the three magnetic materials, calling for flatter magnets for alnico VI, and especially alnico XII, than for alnico V.

Figure 12 illustrates the optimum operating magnetic induction $\mathbf{B}_{op}=\mathbf{B}_r/\lambda_m$ and the corresponding value $\mathbf{B}_{or}$ at the origin of the co-ordinated minor loop, as discussed previously. The optimum magnetic induction $\mathbf{B}_{op}$ for $k=1$, corresponds in all charts to the case of no demagnetization effect, and is equal to the value $\mathbf{B}_o$ of the point of maximum-energy product of the main demagnetization curve.

For the case of the demagnetization by externally applied magnetomotive forces, Figures 13, 14, and 15 exhibit the values v and λ as functions of s/s_o for demagnetization factors c between 1 and 10. In this representation, the per-unit volume v is defined, in accordance with equation 35

$$v=\lambda^2(s/s_o)/B_rH_c \qquad (37)$$

The minimum positions v_{op} of the curves v, and the corresponding values λ_m of the λ functions are marked; the co-ordinated values s_m/s_o, to be utilized in equations 19, 20, and 21, are given by the abscissas. Well-defined minimums are indicated in all cases.

In the final charts, Figures 16 and 17, the per-unit dimensions to be used in connection with equations 30, 34, and 35, and the values $\mathbf{B}_{op}$ and $\mathbf{B}_{or}$ are again computed, permitting the designer similar interpretations, as discussed in connection with Figures 11 and 12. It is an interesting result that the v_{op} curves of the three materials for case B do not intersect. This indicates that for the considered range of c, alnico VI remains always larger than alnico V, and alnico XII always larger than both alnico V and VI.

Although for both cases of demagnetization, the final charts, Figures 11 and 17, contain all basic design information for the optimum design, the design charts, Figures 8, 9, and 10, and 13, 14, and 15, containing the functions v and λ may be important for practical design work, as they demonstrate the effects of deviations from strict optimum designs upon design economy, and sometimes permit the selection of a magnet which has a more suitable shape than the optimum magnet, without noticeably increasing the volume.

Simplified Approximations for Alnico V and VI

The optimum operating magnetic induction $\mathbf{B}_{op}$ for alnico V, in the entire range between $k=1$ and $k=10$ ($c=0$ and $c=10$, respectively), and for alnico VI in the range between $k=1$ and $k=5$ ($c=0$

and $c=5$, respectively), deviates relatively little from the value $\mathbf{B}_o$ of the point of maximum-energy product, as Figures 12 and 17 indicate. This suggests the investigation of the feasibility of an approximation with a constant value of $\mathbf{B}_{op}\approx\mathbf{B}_o$.

The deviations from the minimum volume for $\mathbf{B}_{op}\approx\mathbf{B}_o$, equivalent to $\lambda_m=\mathbf{B}_r/\mathbf{B}_o$ = constant, may be judged from Figures 8, 9, 13, and 14 by the points marked with crosses. It is evident that the deviations do not exceed approximately 5 per cent, for alnico V, and approximately 8 per cent for alnico VI, which justifies the making of the assumption $\mathbf{B}_{op}\approx\mathbf{B}_o$, for preliminary design calculations.

For alnico V and VI for both types of demagnetization, the optimum cross section of the magnet is then given by the equation

$$A_m=\Phi_t/\mathbf{B}_o \qquad (38)$$

For alnico V

$1\leq k\leq 10$
$0\leq c\leq 10$
$\mathbf{B}_o=9{,}680$

For alnico VI

$1\leq k\leq 5$
$0\leq c\leq 5$
$\mathbf{B}_o=7{,}200$

The results for the lengths and volume are different for alnico V and VI, as well as for each of the two cases of demagnetization. They may be derived by replacing the functions which are obtained for the cross-marked points for the z_m and s_m/s_0 values respectively for the variable k and c respectively by straight-line approximations. The results for alnico V are as follows:

Demagnetization by varying external reluctance

$$l_m\approx(0.62k+0.38)\,\mathfrak{F}_t/\mathbf{H}_o \qquad (39)$$

$$V_m\approx(0.62k+0.38)\Phi_t\,\mathfrak{F}_t/(\mathbf{B}_o\mathbf{H}_o) \qquad (40)$$

where $1\leq k\leq 10$.

Demagnetization by externally applied magnetomotive force

$$l_m\approx(0.82c+1)\,\mathfrak{F}_t/\mathbf{H}_o \qquad (41)$$

$$V_m\approx(0.82c+1)\Phi_t\mathfrak{F}_t/(\mathbf{B}_o\mathbf{H}_o) \qquad (42)$$

where

$0\leq c\leq 10$

$\mathbf{H}_o=465$ oersteds
$\mathbf{B}_o=9{,}680$ gausses, in both cases

For alnico VI the results are:

Demagnetization by varying external reluctance

$$l_m\approx(0.46k+0.54)\,\mathfrak{F}_t/\mathbf{H}_o \qquad (43)$$

$$V_m\approx(0.46k+0.54)\Phi_t\,\mathfrak{F}_t/(\mathbf{B}_o\mathbf{H}_o) \qquad (44)$$

where $1\leq k\leq 5$.

Demagnetization by varying externally applied magnetomotive force

$$l_m\approx(0.71c+1)\,\mathfrak{F}_t/\mathbf{H}_o \qquad (45)$$

$$V_m\approx(0.71c+1)\Phi_t\mathfrak{F}_t/(\mathbf{B}_o\mathbf{H}_o) \qquad (46)$$

where

$0\leq c\leq 5$
$\mathbf{H}_o=515$ oersteds
$\mathbf{B}_o=7{,}200$ gausses, in both cases

From the approximate equations it can be derived for case A that for $k=4.6$ alnico V and VI result in magnets of same size, which is in agreement with the more accurate results of Figure 11. No c values can be found which render the size of alnico V and VI equal for case B, which agrees with the results of Figure 12.

Interrelations Between Two Cases of Demagnetization

It is evident that a demagnetization affected by reluctance variations permits also the application of a certain maximum external demagnetizing-magnetomotive force to the circuit, when operating at its normal reluctance $\mathfrak{R}_t$, without causing further demagnetization. Similarly, a magnet, demagnetized by external magnetomotive force, allows a certain maximum open-circuit reluctance.

These interrelations between the demagnetization degrees k and c are often of practical interest. For instance, in permanent-magnet generator design it is important to know which short-circuit armature reaction will be permissible if the rotor is designed for a certain value of k, corresponding to the rotor out of its stator condition.

Equations 52 and 53 in Appendix III can be easily derived from the geometrical conditions existing in Figure 7. $\mathfrak{R}_t$ is thereby in both cases of demagnetization considered a constant value. The results for optimum designs of both demagnetization cases are inserted as dash-dotted curves in Figures 12 and 17, and permit the determination of that value of c which corresponds to a given value of k, Figure 12, and vice versa, Figure 17.

It is obvious that the value of c equivalent to a value of k is always smaller than k. The difference between k and c is thereby influenced by the characteristics of the material.

Final Remarks

The presented design charts and their evaluation are demonstrative examples for several permanent-magnet materials, based on the particular magnetic characteristics as listed before. The given analytical relations enable the designer to develop similar design charts for the same types of material, but having different magnetic characteristics, or for any other material, or for any other desired range of demagnetization. By using the actual magnetic characteristics of the materials, the design accuracy may approach the limits which are given by the tolerances of the material.

Appendix I

Equation for λ, Case A

Using Figure 5, the abscissa values $\mathfrak{F}_2$ and $\mathfrak{F}_3$ of the intersection points 2, between the straight lines $\mathfrak{R}_o$ and m_s, and 3, between the straight line $\mathfrak{R}_o$ and the main demagnetization curve M_{s0}, can be calculated as functions of the shape factor s and parameters k. After simplifications and introduction of the terms

$$z=(s/s_o)/k \qquad (23)$$

$$b=\mu_a\mathbf{H}_c/\mathbf{B}_r \qquad (47)$$

μ_a is always a negative value, therefore b is always negative. The final relation for $\lambda=\mathfrak{F}_r/\mathfrak{F}_2$ reads

$$\lambda=\frac{z-b/k}{2z(z-b)}\left(1+z+\sqrt{(1+z)^2-4az}\right) \qquad (48)$$

Equation for λ, Case B

Using Figure 7, the abscissa values $\mathfrak{F}_2$ and $\mathfrak{F}_3$ of the intersection points 2, between the straight lines $\mathfrak{R}_{ta}$ and m_s, and 3, between the shifting auxiliary line $\mathfrak{R}_a$, connecting points 2 and 0, and the main demagnetization curve M_{s0}, can be calculated as functions of the shape factor s and parameters c. After simplifications and introduction of the term b, according to equation 47 the final relation for $\lambda=\mathfrak{F}_2/\mathfrak{F}_3$ reads

$$\lambda=\frac{1}{2\left(\frac{s}{s_o}-b\right)}\left\{c(1+b)+\left(1-b\frac{s_o}{s}\right)\left(1+\frac{s}{s_o}\right)+\sqrt{\left[c(1+b)+\left(1-b\frac{s_o}{s}\right)\left(1+\frac{s}{s_o}\right)\right]^2-4ab\left[c+\left(1-b\frac{s_o}{s}\right)\right]\left[c-\left(1-\frac{s}{bs_o}\right)\right]}\right\} \qquad (49)$$

Appendix II

Relation for $\mathbf{B}_{or}$, Case A

Using Figure 5, the value $\mathbf{B}_{or}$ can be derived from the ordinate value of point 2, as the intersection of the optimum minor loops with the line of the open-circuit reluctance $\mathfrak{R}_o$. The equation of the optimum minor loop is thereby determined by point 1 and the slope μ_{as} which corresponds to the optimum shape factor s_m. The result is then

$$\mathbf{B}_{0r}=\frac{\mathbf{B}_r}{\lambda_m}\,\frac{1/k-z_m/b}{1-z_m/b} \tag{50}$$

Relation for $\mathbf{B}_{or}$, Case B

Using Figure 7, the value $\mathbf{B}_{or}$ can be found with the same method except that the line of the open-circuit reluctance must be replaced by the line $\mathcal{R}_{ta}$. The result yields

$$\mathbf{B}_{or}=\frac{B_r}{\lambda_m}\,\frac{1-c-(s_m/s_0)/b}{1-(s_m/s_0)/b} \tag{51}$$

Appendix III

Interrelations Between k and c for Optimum Designs

Using Figure 7, the interrelations between k and c can be derived from the reluctance value which corresponds to the shifting auxiliary line $\mathcal{R}_a$ for point 2 of the optimum magnet. It is then $k=\mathcal{R}_a/\mathcal{R}_t$ for the corresponding value of c. The result is

$$k=\frac{1+c-(s_0/s_m)b}{1+(s_0/s_m)b(c-1)} \tag{52}$$

$$c=\frac{z_m-b/k}{z_m-b}(k-1) \tag{53}$$

(By replacing s_0/s_m by s_0/s and z_m by z, the relations can be used also for magnets not representing optimum designs.)

Optimum designs obtained for specified values of k or c do not necessarily represent optimum designs for their interrelated c and k values.

References

1. THE OUTPUT AND OPTIMUM DESIGN OF PERMANENT MAGNETS SUBJECTED TO DEMAGNETIZING FORCES, **A. J. Hornfeck, R. F. Edgar.** *AIEE Transactions*, volume 59, 1940, pages 1017–24.

2. THE ECONOMIC UTILIZATION OF MODERN PERMANENT MAGNETS, **D. J. Desmond.** *Journal*, Institution of Electrical Engineers, London, England, volume 92, 1945, pages 229–52.

3. PERMANENT MAGNETS, **R. L. Sanford.** *Circular C448*, National Bureau of Standards, Washington, D. C., August 10, 1944.

4. DESIGNING STABILIZED PERMANENT MAGNETS, **E. M. Underhill.** *Electronics*, New York, N. Y., volume 17, January 1944, page 118.

5. STABILIZED PERMANENT MAGNETS, **P. P. Cioffi.** *AIEE Transactions*, volume 67, part II, 1948, pages 1540–43.

6. PERMANENT MAGNETS AND THE RELATIONS OF THEIR PROPERTIES TO THE CONSTITUTION OF MAGNETIC STEELS, **E. A. Watson.** *Journal* Institution of Electrical Engineers, London, England, volume 61, 1923, pages 641–60.

7. HANDBUCH DES MAGNETISMS (book), **J. Lamont.** Voss Verlag, Leipzig, Germany, 1867.

8. ABRISS DER DAUERMAGNETKUNDE (book), **J. Fischer.** Springer Verlag, Berlin-Goettingen-Heidelberg, Germany, 1949, page 189.

◆

Discussion

P. P. Cioffi (Bell Telephone Laboratories, Murray Hill, N. J.): I have also been concerned with the problem of optimum design of permanent magnets when stabilized against demagnetizing forces.[1] My paper discussed three aspects of the problem: Optimum operation under conditions of (a) external variable reluctance; (b) applied external demagnetizing force; and (c) the interrelationship between the two methods of stabilization. It was shown that for either method the state of magnetization, as represented by the terminus of the tip of the minor loop on the air-gap line, is dependent upon the reversible permeability, the degree of demagnetization, and the slope of the air-gap line. The locus of this point, as the air-gap line slope changes from zero to infinity, is a curve passing through $\mathbf{H}_c$ and $\mathbf{B}_r$. This curve, which I called the secondary demagnetization curve, can be used with its associated energy-product curve identically as the main demagnetization and energy-product curves are used in design problems. The point of intersection of this curve with the air-gap line defines the state of magnetization of the permanent-magnet material for the given degree of stabilization. Analytical and graphical methods were developed for determining the secondary demagnetization curves for the two methods of stabilization, and families of such curves with their associated energy-product curves were presented for the four most important permanent-magnet materials for various degrees of stabilization.

The present paper by H. K. Ziegler appears to present identically the same aspects of the problem as the three mentioned. The novelty of the method rests chiefly on the use of new terms and nomenclature introduced by the author such as "ideal magnet," "magnification factor," and "shape factor ratio." The nature of the proposed method obscures the physical significance of the problem and yields results which have no advantage in application.

REFERENCE

1. See reference 5 of the paper.

George C. Uchrin (Signal Corps Engineering Laboratories, Fort Monmouth, N. J.): The paper presented has brought out the results that the proper choice of the slope of the minor hysteresis loops is of vital importance for the design economy. This fact can be seen still more clearly if, for instance, in the simplified relation for alnico V, as given by equation 40 of the paper, the value of the average slope μ_a is introduced as a design parameter, instead of the numerical value which applies for that particular chosen value of $\mu_a=-5.0$.

The outcome of such an investigation with variable values of μ_a has given the following relation for the optimum magnet volume v_{op} per-unit external magnetic energy, for alnico V

$$v_{op}=\frac{k}{\mathbf{B}_o\mathbf{H}_o}\left[1-0.076\left(\frac{k-1}{k}\right)|\mu_a|\right]\ \frac{\text{centimeters}^3}{\text{maxwell-gilberts}}$$

where

k = demagnetization degree $\mathcal{R}_o/\mathcal{R}_t$ for demagnetization by reluctance increase from $\mathcal{R}_t$ to $\mathcal{R}_o$

$\mathbf{B}_o\mathbf{H}_o$ = maximum external energy product of main demagnetization curve

$|\mu_a|$ = absolute value of average slope of straight-line replacement for minor hysteresis loops, in the zone to which optimum designs are confined

This equation was found to be correct within approximately 1 per cent of the values of the charts. This equation contains, as a design parameter, the average value μ_a of the slope of the minor hysteresis loops. Also contained are the values of the maximum energy product $\mathbf{B}_o\mathbf{H}_o$, and a factor 0.076 which depends upon the fullness factor of the main demagnetization curve. This confirms the fact that for permanent magnets subjected to demagnetizing effects, the slope of the minor hysteresis loops is an important factor, in addition to the maximum-energy product, for judging the over-all qualities of a permanent-magnet material. It is evident that materials having a large slope μ_a in the zone which applies to optimum designs offer smaller magnets than materials with smaller slopes. This is well understood, because a material with a steep slope of the minor loops gives a better recovery after demagnetization. For a general conclusion that materials with a large slope are more desirable than those with a small slope, it must be considered that for some applications the slope of the minor loops also influences the final operating characteristic of the magnetic circuit of which the permanent magnet is a part.

Hans K. Ziegler: Mr. Cioffi's paper has been given recognition by reference 5 of the author's paper. In reply to his discussion, it is necessary to compare both papers in order that important differences and results of the presented methods can be clearly understood:

In concurrence with reference 4 of his paper Mr. Cioffi advocates, for all practical design purposes, the assumption of a constant slope of the minor hysteresis loops. As he does not specify any value for such constant slope, it must be assumed that he also agrees with reference 4 in choosing a value equal to the differential permeability at the point $\mathbf{B}_r$. His paper includes the grade of alnico V.

The author's paper has brought out, as a new result, the practical significance of the variation of the slope of the minor hysteresis loops and of the proper choice of an average constant value upon the design economy. Alnico V, as a material with very large slope variations, was thereby singled out to expose the consequences originating from the previous lack of information regarding the significance of these factors. It is deemed by the author that weight savings for alnico V, of the order of 25 per cent, for frequently occurring practical demagnetization factors of $k=5$, obtainable through the utilization of the new results, represent a worth-while improvement. Of course, for materials with small slope variations, the savings will be smaller or even negligible, but it should be kept in mind that presently alnico V is the most important and widely used alloy.

In view of the fact that the author's paper employs an unprecedented method of attack to the problem, it is a logical consequence that certain new terminology would be introduced. The author does not agree to the inference of Mr. Cioffi's discussion that such new terminology disguises the physical significance of the subject involved. The design method is a step-by-step logical engineering procedure which is as feasible for the designer as the application of maximum-energy product considerations from a physical point of view. Designers have been using the method successfully for the past 6 years without difficulties. Regardless of the merits of the one or the other design procedure, it has always been considered a healthy research policy to exploit all possible means of attack for the solution of a problem in a relatively new field. As a rule, some new results associate themselves with each new approach. The uncovering of the presented new relations between the minor hysteresis loops and the design economy are an example of such associated results. It is an interesting coincidence that Mr. Cioffi's and the author's methods have been developed completely independent of each other. Mr. Cioffi's paper was presented in October 1948, while the author's method was filed with the Signal Corps Engineering Laboratories in June 1947, and was released only recently.

With reference to the working efforts for the execution of either method for the development of design charts, it is seen that two basic curves are needed for both. In Mr. Cioffi's paper these are the secondary demagnetization and the related energy product curves, whereas in the author's paper these are the magnification factor and the related optimum volume curves. In Mr. Cioffi's method a combined graphical and analytical procedure is employed for deriving the secondary demagnetization curve, utilizing a graph of a main demagnetization curve to select pairs of $\mathbf{B}_1$ and $\mathbf{H}_1$ values. The author uses a completely analytical procedure for the calculation of the magnification-factor curve, whereby the analytical relation of the main demagnetization curve is incorporated. This accounts for the more complicated analytical expression. The further evaluation of the curves and the necessary efforts are, for both methods, quite similar. In addition to the method for the development of design charts, the author's paper has produced simple analytical relations for the direct precalculation of optimum magnet dimensions for alnico V and VI. They represent satisfactory approximations for a wide range of practical applications, and have proved to be a very expeditious means for designing permanent-magnet apparatus.

It should be understood that Mr. Cioffi's method of design, when modified by using the actual μ_r values associated with each selected pair of values $\mathbf{B}_1$ and $\mathbf{H}_1$, or by the introduction of proper average values of the slope of minor hysteresis loops, as emphasized in the author's paper, will eventually lead to results which are identical with those obtained from the author's method. It will be up to the designer to decide which method offers best aid for his particular design problems.

The relation quoted by Mr. Uchrin is gratefully acknowledged as a contribution to further support the significance of the proper choice of the average-slope value of the minor hysteresis loops on the design economy.

Electronic Frequency Changer Used as Nonsynchronous Tie Between A-C Power Systems

HAROLD WINOGRAD
FELLOW AIEE

EARLY in 1943 a 6,667-kw electronic frequency changer, using mercury-arc rectifiers, was installed in the Gary Sheet and Tin Mill of the United States Steel Corporation, at Gary, Ind., to provide a nonsynchronous tie between the 25-cycle and 60-cycle power systems, for the controlled interchange of power in either direction.[1] All the 25-cycle power is generated in the steel company's plants in the Chicago district. The 60-cycle system is supplied by the steel company's generating equipment and by the Northern Indiana Public Service Company. The generating capacity of each system is over 200,000 kw.

The interconnection had to be made through a nonsynchronous tie because of the frequency variations on the 25-cycle system. The electronic-type frequency changer was selected by the engineers of the United States Steel Corporation after investigating several alternative methods, including the Scherbius system, two a-c-to-d-c motor-generator sets, and two rotary converters. Preference was given to the mercury-arc frequency changer, in spite of its higher first cost and the pioneering involved, for the following reasons:

1. It consists of stationary equipment and requires no special foundation.
2. There are no slip rings, commutators, and brushes, and the maintenance cost was expected to be lower.
3. It is easy to start without any shock on the system and it requires no synchronizing. It does not feed into a fault on the a-c system.
4. The power flow can be controlled easily.
5. It has a higher all-day efficiency.

In the original installation, multianode-type mercury-arc rectifiers were employed. In 1948 these were replaced by single-anode type excitron rectifiers and the rating was increased to 8,000 kw. A second 8,000 kw unit was installed in 1949, bringing the capacity of the installation to 16,000 kw, with a 2-hour rating of 20,000 kw.

In the electronic frequency changer a d-c link serves as the tie between the two power systems. A power rectifier converts power from one a-c system to d-c power, which is in turn converted to a-c power into the other system by a power inverter.

Before describing the Gary frequency-changer installation, the operating principles of a power inverter and an electronic frequency changer will be explained. The term "rectifier unit" is applied to the physical equipment consisting of a rectifier transformer, rectifier, and associated auxiliaries and switching equipment. The terms "power rectifier" and "power inverter" are applied to the power-conversion functions of a rectifier unit.[2]

Operation of Power Inverter

BASIC PRINCIPLES

The operation of a power inverter is basically the same as that of a power rectifier, and a rectifier unit can be used alternately for either conversion function. The rectifier tubes are connected to a multiphase winding of a rectifier transformer, and conduct the direct current of the unit in sequence during each cycle. The current can flow through the tubes in one direction only, from anode to cathode.

Paper 53-53, recommended by the AIEE Electronic Power Converters Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted November 7, 1952; made available for printing November 20, 1952.

HAROLD WINOGRAD is with the Allis-Chalmers Manufacturing Company, Milwaukee, Wis.

The author gratefully acknowledges the valuable contributions made to this development by R. W. Holman, United States Steel Corporation, and by E. Bender, I. K. Dortort, J. Kiefer, and A. D. Millikin, present and former associates of the author.

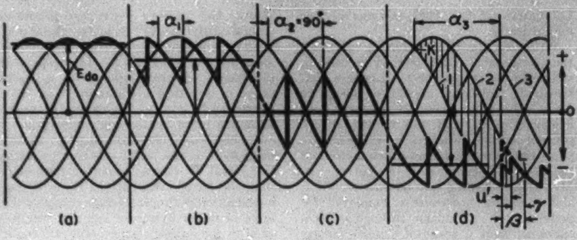


Figure 1. Transition from power rectification to power inversion by phase control

The essential difference between the two functions is that in a power rectifier the tubes conduct during the positive half-cycles of their phase voltages, and the average direct voltage is positive and in the same direction as the current; in a power inverter the tubes conduct during the negative half-cycles of their voltages, so that the average direct voltage of the unit is negative and opposite to the direction of the current.

The firing time in the cycle is controlled by the control grids of the tubes. A tube is permitted to fire when its grid voltage with respect to the cathode is changed from negative to positive. The control of the firing point in the cycle, called "phase control," is used also for regulating the magnitude of the direct voltage.

The operation of phase control and the transition from power rectification to inversion are illustrated in Figure 1. The sine waves represent the phase voltages of a 6-phase star-connected rectifier unit, as shown in Figure 2. In section (a) of Figure 1, the tubes are permitted to fire when their phase voltages are highest. Each tube conducts for 1/6 of a cycle (neglecting the commutating period), and the direct voltage, indicated in heavy outline, has a maximum value. In section (b) the firing point of each tube in its voltage cycle is retarded by the phase-control angle α_1, and the average value of the direct voltage is reduced. In section (c), with a 90-degree angle of retard, the average direct voltage is zero. It is assumed that the d-c circuit has sufficient inductance to maintain the flow of direct current. In section (d) the firing of the tubes has been retarded by an angle α_3, so that they conduct during the negative half-cycles of the phase voltages, and the average direct voltage is negative. The direct current continues to flow in the same direction, and the reversal of the voltage represents a reversal of power flow, which was effected entirely by phase control.

To permit the flow of direct current through a rectifier unit in the opposite direction from its voltage, an external direct voltage has to be applied in the same direction as the current and of sufficient magnitude to overcome the negative voltage. The external voltage could be supplied by another rectifier unit which might be connected to another a-c system.

Such an arrangement is shown in Figure 2, in which are indicated the directions of the current, voltages, and power. This is the basic circuit of an electronic frequency changer. Either unit may be operated as a power rectifier or a power inverter, and the direction of power flow can be changed by phase control without interrupting the d-c circuit. The magnitude of the direct current and the transmitted power is determined by the relative values of the direct voltages, which can be regulated also by phase control. The negative direct voltage of a power inverter is comparable to the back electromotive force of a d-c motor.

The operation of a power inverter depends on the ability to delay the conducting period of the tubes and to commutate (transfer) the direct current between tubes of successive phases. Commutation is possible only when the incoming phase has a higher voltage than the outgoing phase. Referring to Figure 1, this condition for a pair of successive phases prevails between the intersection point K of the positive half-cycles and the intersection point L of the negative half-cycles of their voltages. If, for example, the current is not commutated from phase 1 to phase 2 before point L is reached, phase 1 will continue to conduct beyond L and into its positive half-cycle. The back-voltage of the power inverter would then disappear, and the external d-c power supply would be short-circuited through phase 1 of the inverter transformer and the tube connected to it. Point L is the commutation-limit point, and commutation must be completed before it is reached.

Operating Requirements

In Figure 3 are shown the circuit relations of a power inverter with a 6-phase double-wye connection which is the most generally used rectifier connection and is employed for the Gary frequency changer. It consists of two 3-phase commutating groups paralleled through an interphase transformer. Each group conducts the direct current I_c, which is one-half of the total current I_d. Because of the phase displacement between the groups, the direct voltage of the unit has a 6-phase wave shape.[3]

At the right of Figure 3 are shown the phase-voltage sine waves of one commutating group. During each cycle, the direct current I_c is conducted in sequence by the three phases and their tubes. Each tube is fired by its control grid at a point on its voltage wave advanced by the angle β from the commutation-limit point L. The current is commutated from the outgoing to the incoming phase during an angle u', and the commutation is completed short of point L by the angle

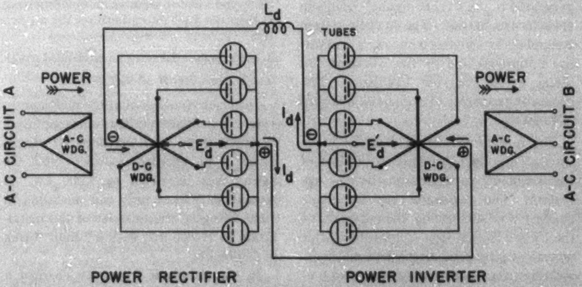


Figure 2. Elementary circuit of electronic frequency changer, showing directions of direct current, direct voltages, and power

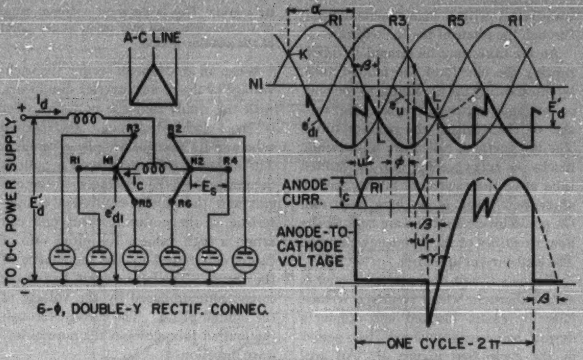


Figure 3. Current and voltage relations of power inverter with 6-phase double-wye connection

γ, which is called the "margin angle." The direct voltage e_{d1}' of the commutating group, shown in heavy outline, is equal to the voltage of the conducting phase. During the commutating period it is equal to the average of the overlapping phase voltages e_u. The average direct voltage of a group E_d' is also the average voltage of the unit, since the two groups are paralleled.

The duty on a tube of a power inverter is indicated by its anode-to-cathode voltage, as shown in Figure 3 for the tube of phase $R1$. It is obtained by taking the difference between the voltage of phase $R1$ and the cathode voltage e_{d1}'. During the conducting period of the tube, the voltage is equal to the arc drop and is low. It is negative during the margin angle γ, and positive beyond point L until the start of the next conducting period. It attains peak positive values equal to the peak value of the phase-to-phase voltage. The margin angle is needed for deionization of the control grid so that it will regain control and block the anode during the positive period. The commutating angle u' increases as the load current is increased, and the angle of advance β has to be sufficiently large to provide the needed margin angle for the highest current to be commutated.[4]

It should be noted in Figure 3 that the center of the anode current of $R1$ is advanced by an angle ϕ from the center of the negative half-cycle of the phase voltage, as a consequence of requiring an angle of advance. The significance of this is that the fundamental component of the line alternating current, which is a reflection of the anode currents, leads the output line-to-neutral voltage of the power inverter (which is equal and opposite to the applied voltage). The a-c system must therefore be able to receive current of leading power factor. In practice this means that the system should have sufficient synchronous machine capacity, or static capacitors, capable of supplying lagging kilovars to compensate for the leading kilovars put out by the power inverter, in addition to the kilovars required by the load.

On the basis of the preceding discussions, there are two system requirements for the operation of a power inverter:

1. The system must provide the alternating voltages which are required for producing the negative direct voltage of the inverter and for commutating the direct current between tubes. The required voltages can be provided by a-c generators or a synchronous condenser.

2. The system must be capable of supplying the lagging kilovars required by the load and the power inverter. This can be provided by a-c generators, a synchronous condenser, or static capacitors.

Both of these requirements were satisfied for the Gary frequency-changer installation by existing system conditions.

Regulation of Power Flow

The flow of power through an electronic frequency changer may be regulated by phase control of the output voltage of the power rectifier, or the back-voltage of the power inverter, or both. On the Gary frequency changer, the power inverter operates at a fixed angle of advance, and the power flow is controlled by regulating the voltage of the power rectifier only. This method is illustrated by the voltage regulation characteristics of Figure 4.

The terminal direct voltage of a power rectifier is given by the expression[2]

$$E_d = E_{d0} \cos \alpha - (E_x + E_r + E_a) \qquad (1)$$

where

E_{d0} is the theoretical no load voltage, and is equal to $1.17E_s$ for the circuit of Figure 3

E_s = phase voltage of rectifier transformer d-c winding

α = phase-control angle of retard

E_x = voltage drop due to commutating reactance

E_r = voltage drop due to resistance losses in rectifier transformer and any other elements of the unit

E_a = arc-drop voltage of rectifier

The voltage drops E_x and E_r are directly proportional to the load current, while E_a is nearly constant. The voltage regulation characteristic is therefore a straight line sloping downward at increasing load. The maximum voltage is obtained when α is zero, as shown by line E_{dm} in Figure 4. For other values of α, regulation lines are obtained, such as E_{d1} and E_{d2}, which are parallel to E_{dm}.

For a power inverter, the terminal direct voltage is the voltage required to overcome the theoretical back-voltage, plus the voltage drops produced by the

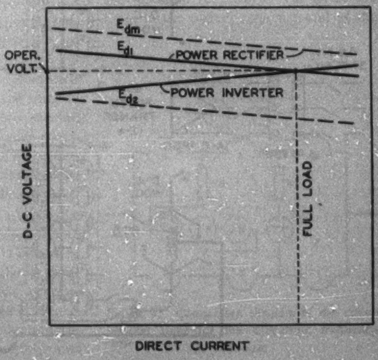


Figure 4. Voltage regulation characteristics of power rectifier and power inverter, illustrating load control of electronic frequency changer

flow of current, and is given by the expression

$$E_d' = E_{d0} \cos \beta + (E_x + E_r + E_a) \qquad (2)$$

in which β is the angle of advance. The voltage regulation characteristic is therefore a straight line sloping upward at increasing current, as shown in Figure 4. The position of this line is determined by the value of β.

Since the terminal voltages of the power rectifier and the power inverter of the frequency changer must be equal, the direct current is determined by the intersection point of their voltage-regulation characteristics. This point can be shifted to control the direct current and the transmitted power by varying the phase-control angle α, which shifts the voltage regulation line of the power rectifier. The inherent voltage regulation of the Gary frequency changer, between full load and light load, is about 5 per cent.

OPERATING FAULTS

The anode-to-cathode voltage of a power rectifier is similar to that shown in Figure 3, when viewed upside down. It is negative during the greater part of the nonconducting period in the cycle. The major fault of a power rectifier is an arc-back, which results when a cathode spot develops on an anode during the inverse period. Reverse current flows to the faulted tube from the tubes connected to the other phases, resulting in an unbalanced short-circuit on the rectifier transformer with a high reverse current through the phase of the faulted tube.

An arc-back may be cleared by interrupting the circuit of the faulted tube, by interrupting the a-c supply to the unit, or by applying negative blocking voltage to the grids of the rectifier tubes. The last method, called arc suppression, blocks the tubes feeding into the faulted tube after their currents fall to zero in sequence. There is no back-feed from the power inverter into an arc-back on the power rectifier of the frequency changer. The load current drops off.

The major fault of a power inverter is an arc-through, which results if a tube fails to remain blocked and conducts during the positive nonconducting period in the cycle (see Figure 3). An arc-through may be caused by one of the following conditions:

1. When the current is not commutated away from the tube before the commutation-limit point L, or if the margin angle is too small for deionization of the grid.
2. Interruption or single-phasing of the a-c system voltage, so that the back-voltage of the power inverter disappears or is noncontinuous.
3. A fault in the grid-control circuit.
4. Failure to commutate away the current of a tube because of a misfire on the tube of the succeeding phase.
5. An arc-back on a tube during the margin angle, when the anode is negative to the cathode, may cause the tube to conduct when the anode voltage reverses.
6. Random and transient loss of control by a grid. This is the most common cause of arc-throughs, and may occur at any point of the positive period.

When an arc-through occurs, the back-voltage of the power inverter disappears, with the result that the power rectifier is short-circuited through the faulted tube and its transformer phase winding. The ultimate current of such a fault is determined by the resistance of the circuit and the voltage-regulation characteristic of the power rectifier, and can attain a high value. Its initial rate of rise is determined by the inductance of the circuit. The fault current has an a-c component, produced by the voltage of the faulted phase. The current through the faulted tube flows in the normal forward direction.

An arc-through can be cleared by removing the d-c power supply. This may be effected by one of the following methods:

1. Interrupting the d-c circuit.
2. Applying arc suppression to the power rectifier by grid blocking.
3. Interrupting the a-c supply to the power rectifier.

An arc-through produced by a random nonsustained cause may clear itself by recommutation. The tube of the succeeding phase, fired in its normal sequence, can commutate the current away from the faulted tube, if the current attained at that point is sufficiently low so that the commutating angle is smaller than the

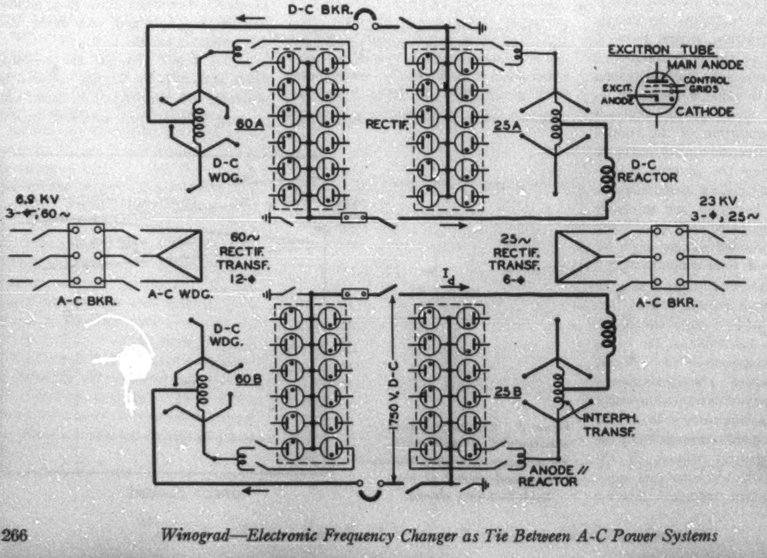


Figure 5. Power-circuit diagram of 8,000-kw electronic frequency-changer unit installed at Gary

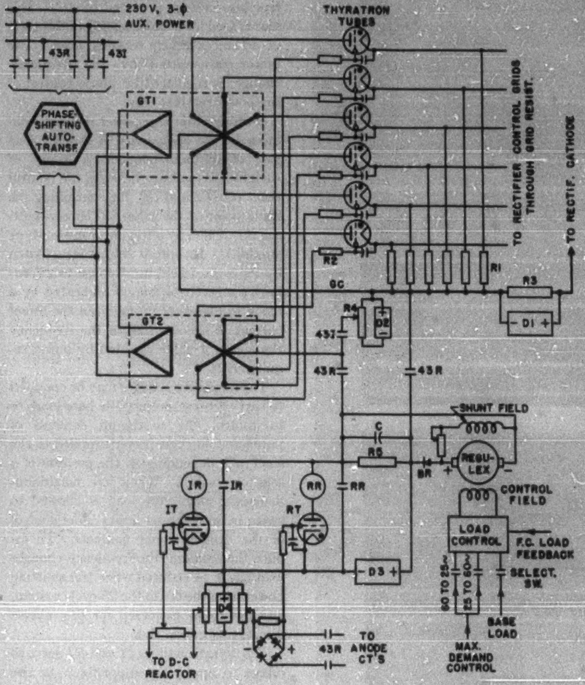


Figure 6. Grid-control and load-regulating circuit of Gary frequency changer

angle of advance. This could happen if the fault occurs late in the positive period. The same result can be obtained for arc-throughs occurring at any point of the positive period by inserting into the d-c circuit sufficient inductance to slow up the rate of rise of the current, so that it does not reach the commutation-limit value at the firing point of the succeeding tube. The d-c inductance is also beneficial in keeping down the magnitude of the current resulting from sustained faults, which are cleared by one of the other methods.

The several protective measures discussed, including d-c reactors, are incorporated in the protection system of the Gary frequency changer.

Gary Frequency Changer

POWER CIRCUIT

The power circuit diagram of one unit of the Gary frequency changer installation is shown in Figure 5. The a-c windings of the 25-cycle and 60-cycle rectifier transformers are connected to their respective systems through an oil circuit breaker.

The unit consists of two 4,000-kw sections, each comprising a 25-cycle and a 60-cycle rectifier group, which are connected in series on the d-c side. One group operates as a power rectifier, the other as a power inverter; their functions depend on the desired direction of power flow. Each group consists of a 12-tube pumped excitron rectifier connected to a 6-phase d-c winding of the rectifier transformer. A pair of tubes is connected to each phase through an anode-paralleling reactor, which consists of a circular-wound iron core, through which the anode cables are looped in opposite directions.

The d-c circuit operates at 1,750 volts. It is provided with a d-c breaker and an oil-immersed d-c reactor. Grounding switches are provided for the cathode terminals. Only the grounding switches of the power rectifiers are closed in service, which makes it possible to operate the load-regulating equipment close to ground potential. Disconnecting switches in the d-c and anode circuits permit inspection or servicing the rectifiers of one section while the other section remains in operation.

The d-c reactor, besides suppressing arc-throughs and limiting the fault currents, assists the load regulator by slowing up changes of the load current resulting from variations of the a-c system voltages. It also suppresses the harmonic components in the direct current. The reactors of unit *1* have an inductance of 15 millihenrys. Unit *2* has 12 millihenry reactors which operate on a 10-millihenry tap.

The 6-phase d-c windings of the 60-cycle transformer are zig-zagged to produce a 12-phase combination. The 60-cycle transformer of unit 2 consists of two cores with wye and delta a-c windings, and 6-phase double-wye d-c windings, producing a 12-phase combination displaced by 15 degrees from that of unit *1*. The two units combined therefore constitute a 24-phase system on the 60-cycle side, and the resulting line alternating current has a wave shape closely resembling a sine wave.

On the 25-cycle side, both units have 6-phase transformers, one with a delta-connected, the other with a wye-connected a-c winding, producing a 12-phase combination. A larger number of phases was provided for the 60-cycle side to avoid the possibility of telephone interference on the power system of the Northern Indiana Public Service Company.

GRID-CONTROL AND LOAD-REGULATING CIRCUIT

Figure 6 shows a simplified diagram of the grid-control circuit for each rectifier. A negative blocking voltage is applied to the control grids of the rectifier tubes by a direct-voltage supply *D1*, through resistors *R1*. (Circuit elements *D1* to *D4* are selenium rectifiers.) The grids are fired, in the correct sequence and at the

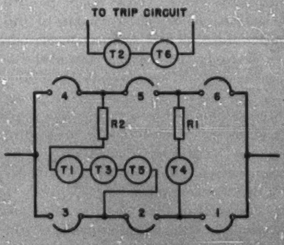


Figure 7. Connection diagram of d-c circuit breaker used for the electronic frequency changer

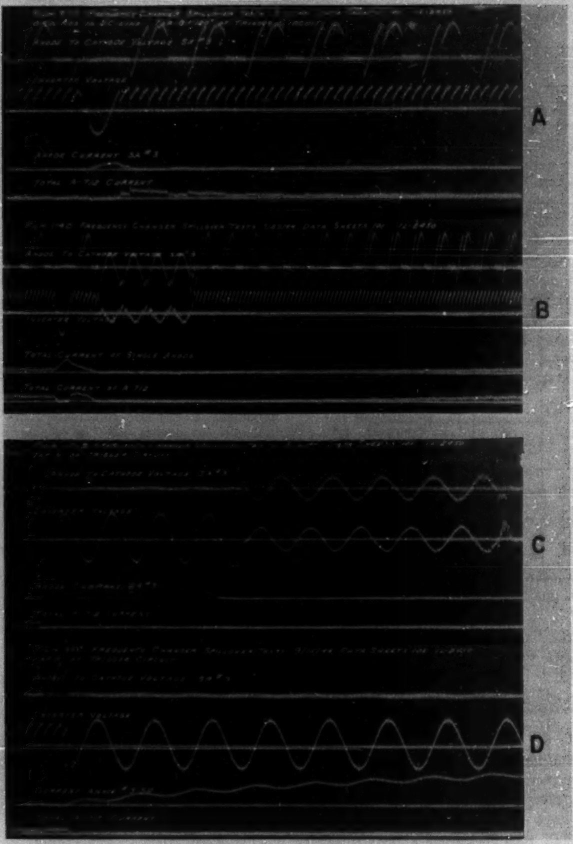


Figure 8. Oscillograms of arc-throughs in power inverter of electronic frequency changer, illustrating several methods of clearing the fault

proper point of the cycle, by means of thyratron tubes. They apply positive voltages from a 6-phase winding of grid transformer *GT1*, of sufficient magnitude to overcome the blocking voltage of *D1*. The resistors *R1* serve as individual loading resistors for the tubes, and the firing voltages appear across them. The average current of each tube is about 0.25 ampere. The connection *GC* is the common reference terminal for the cathodes of the thyratron tubes.

The grids of the thyratron tubes are controlled by the 6-phase winding of grid transformer *GT2* and a negative bias voltage between the neutral point of the grid transformer and terminal *GC*. The firing point of the thyratron tubes, and of the corresponding rectifier tubes, is determined by the phase position of the grid alternating voltages and the magnitude of the bias voltage.

The phase position of the grid-firing voltages in relation to the anode voltages of the rectifier unit, which determines its operation as a power rectifier or a power inverter, is set by a tapped phase-shifting autotransformer, and contactors *43R* and *43I*. The contactors are operated by a selector control switch, which establishes the direction of power flow.

For power inverter operation, a contact of *43I* connects the neutral point of grid transformer *GT2* to *GC* through a negative bias voltage provided by *D2*. This bias is used for a closer setting of the angle of advance, which remains fixed. The power transmitted by the frequency changer is regulated by phase control of the power rectifier only.

For operation as a power rectifier, contacts *43I* are open and *43R* closed. One pair of contacts connects a source of variable-bias voltage between the neutral point of *GT2* and *GC*, for controlling the firing point of the tubes. The bias voltage is supplied by a Regulex rotating-type amplifier. Its output voltage is regulated by its control field in response to a feedback load-control system, indicated by a block in Figure 6, to maintain the direct current in the d-c link of the frequency changer at a value dictated by power requirements.

The frequency changer may be operated to carry a preset adjustable base load, or to control the maximum demand of purchased 60-cycle power supplied to the steel mill, depending on the position of a selector switch. When on maximum-demand control, the load is allowed to swing between preset limits under control of the demand-meter contacts. To reduce the demand, the frequency changer load has to be reduced when transmitting power from the 60- to the 25-cycle system, and has to be increased for the reverse direction of power flow.

The thyratron tubes *IT* and *RT* serve as relays to apply arc suppression to the power rectifier upon occurrence of an arc-through or an arc-back. Tube *IT* is triggered on an arc-through in the power inverter by a voltage from a secondary winding of the d-c reactor. An arc-back in the power rectifier triggers tube *RT* with a voltage pulse from anode current transformers. The firing of either tube applies to the grids of the control tubes a blocking bias voltage from *D3*, which exceeds the maximum value of the grid alternating voltages. The blocking circuit is opened, after a short time delay, by the series relays *IR* and *RR;* their contacts short-circuit the tubes and coils, thereby extinguishing the tubes and resetting the relays. A blocking rectifier *BR* prevents the flow of current from *D3* through the Regulex amplifier.

One set of load-control and blocking equipment is provided for each section of the frequency changer. The selector switches and the demand control are common for both sections.

PROTECTION

The units are provided with overcurrent protection on the a-c side of each rectifier transformer. Overcurrent relays, with

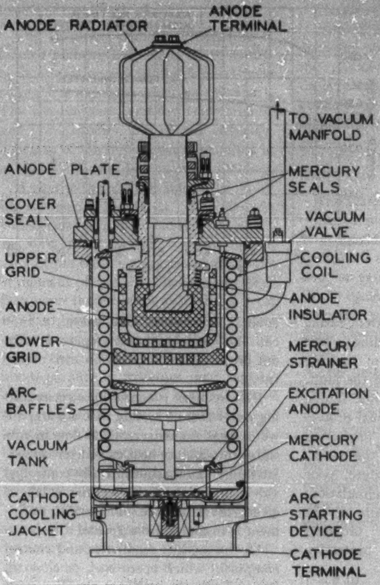


Figure 9. Cross section of excitron rectifier tube used for the Gary frequency changer

time-delay and instantaneous elements, are used for this purpose and are connected to trip the a-c breakers. On an arc-back in the power rectifier, arc-suppression by grid blocking is applied to the faulted rectifier, as previously described. This is backed up by tripping of the a-c breaker. Arc suppression provides fast interruption of the fault current, but cannot be relied on for absolute protection in all cases.

An arc-through in the power inverter is usually suppressed by the d-c reactor, supplemented by application of arc suppression to the power rectifier, as has been described. A sustained arc-through is cleared by the d-c breaker, which is tripped at a preset overcurrent in the d-c circuit. Because of the inductance of the d-c reactor, a standard d-c breaker was not available for interrupting the circuit during a fault. A 6-pole anode breaker,[6] used for other rectifier applications, was adapted for this purpose.

A circuit diagram of the breaker is shown in Figure 7. The six poles are connected in series parallel. The breaker is closed by a motor. Each pole is held closed by an electromagnet (not shown) against the pull of a spring; it is tripped by a d-c trip coil which bucks the flux at the pole faces of the electromagnet and releases the holding armature. The series reverse-current trip coils of the standard anode breaker were replaced by suitable coils designated in Figure 7 by *T1* to *T6*, for the breaker poles of corresponding numbers.

The poles are tripped in sequence, cutting in two steps of current-limiting resistors before final interruption. On an overcurrent in the d-c circuit, tripping current is applied to coils *T2* and *T6* from the station battery. The opening of poles 2 and 6 inserts resistor *R1* into the d-c circuit, in series with trip coil *T4*. The opening of pole 4 inserts resistor *R2*, in series with the trip coils of the remaining poles, which interrupt the d-c circuit. The d-c breaker can be tripped manually by opening the coil circuit of the holding electromagnets.

The oscillograms of Figure 8 illustrate several methods of clearing an arc-through in a power inverter. The tests were made on one section of unit *1*, with the 60-cycle side as the power inverter. The arc-throughs were produced by pulsing the grids of an excitron tube during the positive nonconducting period. For the oscillograms in Figures 8(A), 8(B), and 8(D) the tube was connected in parallel with one anode of an operating rectifier (*A-712* in the oscillograms), but did not itself carry normal load current. For the oscillogram in Figure 8(B) the faulting tube was one of a 6-tube rectifier, which carried about one-half the load current of the section.

The top trace of each oscillogram is the anode-to-cathode voltage of the faulting tube; the second trace is the terminal d-c

Figure 10. View of 60-cycle side of frequency-changer building at Gary Sheet and Tin Mill

Figure 11. Interior view of frequency-changer installation, showing the equipment of unit 2

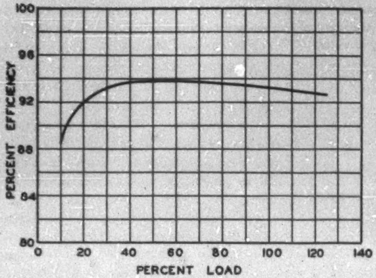


Figure 12 (left). Over-all efficiency of 8,000-kw electronic frequency-changer unit 2

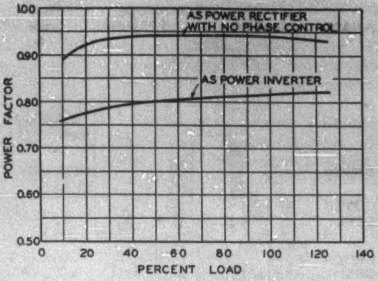


Figure 13 (right). Power factor on the 60-cycle side of 8,000-kw electronic frequency change

voltage of the power inverter (negative above the zero line); the third trace shows the fault current; the bottom trace is the current of the parallel rectifier.

In the oscillogram in Figure 8(A) the arc-through was suppressed by recommutation. Besides the reactor, there was a current-limiting resistor in the d-c circuit. There was no resistor in the tests recorded by the other oscillograms. The anode-to-cathode voltage is equal to the arc drop while the current is flowing, and is close to the zero line. During the arc-through, the inverter voltage is the voltage of the faulted phase, which accounts for the loop in the positive direction, produced by the positive half-cycle of the voltage.

Arc suppression was applied to the power rectifier (25 cycles) in the oscillogram in Figure 8(B), but it took effect only after the arc-through was suppressed by recommutation, due to the influence of the d-c reactor in slowing up the rise of fault current. After the removal of arc suppression, the direct current was gradually restored to normal, as indicated by the lower traces.

The fault current was interrupted by the d-c breaker in the oscillogram in Figure 8(C). The inverter voltage is the voltage of the faulted phase. The voltage traces after the breaker opened are due to the potentiometer connections for the oscillograph elements.

In the oscillogram in Figure 8(D) the fault was cleared by tripping the a-c breaker of the power rectifier. Unfortunately, the opening of the breaker was not recorded on the oscillogram. The peak value of the current near the right end of the trace is 10,100 amperes. The current scale is the same for all the oscillograms. The a-c component of the fault current was produced by the voltage of the faulted phase.

The d-c component of the ultimate fault current resulting from an arc-through was calculated to be about 37,000 amperes, which is over 50 times the current rating of the faulted phase winding. This emphasizes the importance of limiting the current by fast clearing of the fault.

Besides the protection against faults, the rectifier units have the usual protective devices and circuits for protection against overtemperature of the transformer, and for overtemperature, poor vacuum, stoppage of water circulation, and sustained excitation failure of the rectifiers. Annunciators are provided for indicating operation of protective devices.

Rectifiers and Their Auxiliaries

A cross-section view of an excitron tube used for the Gary frequency changer is shown in Figure 9. It has the same general design as the tubes used for other rectifier applications.[6] Its main features are:

1. A d-c excitation arc is maintained continuously at the cathode while in operation, and the firing of the anode is controlled by grids.

2. The active part of the tube is enclosed by a cooling coil, which is insulated from the tank and serves the dual purpose of shielding the tank walls from the arc and conducting the cooling water in series with the cathode jacket. The tank is solidly connected to the cathode, and the tank support is the cathode terminal.

3. The tube has two control grids. This was found necessary for power-inverter service, to provide the required blocking reliability.

The excitation current, of about 10 amperes per tube, is supplied by a selenium rectifier. The excitation arc is started by connecting the excitation anode to the cathode with a jet of mercury, ejected by a nozzle plunger which is pulled down by a solenoid.

The 12 tubes of each rectifier are mounted on a base which is supported on insulators. They are connected to the vacuum pumps by a vacuum manifold through shut-off valves on the individual tubes. By closing these valves, a tube can be removed and replaced without breaking the vacuum on any of the tubes.

Each rectifier is cooled by a recirculating water cooling system with a water-to-water heat exchanger. The flow of raw water through the heat exchanger is regulated by a thermostatically-controlled valve, to maintain the rectifier at about 48 degrees centigrade. The cooling equipment is insulated from ground.

All the rectifier auxiliaries and control equipment which operate at, or close to, the cathode potential are supplied with voltage through an insulating transformer. The excitation and grid-control equipments, the vacuum meter, and other rectifier control devices are mounted in a cubicle, which is also insulated from ground.

Equipment is also provided for preheating the rectifiers with current at low voltage following a prolonged shut-down.

Operating Controls

The operating and load-regulating control equipment of each unit, as well as the indicating and recording instruments, are mounted on a control switchboard.

To put a unit in operation, the selector switches are set for the desired direction of power flow and method of load control. A "start" push button is operated to energize the grid-control and excitation circuits of the four rectifiers. The a-c breakers are then closed by control switches, followed by closing of the d-c breakers. The filaments of the thyratron tubes are energized continuously. The unit is shut down by a "stop" push button, which trips the a-c and d-c breakers.

When the unit is shut down, the load regulator is set automatically to pick up a light load when the unit is restarted. When operating on base-load control, the load of each section can be set by a hand-operated rheostat on the switchboard.

INSTALLATION

The frequency-changer installation is located midway between a 69/23-kv 25-cycle substation and a 132/6.9-kv 60-cycle substation, which are about 200 feet apart. The a-c breakers, rectifier transformers, and d-c reactors are placed outdoors. The rest of the equipment is installed in a 44- by 115-foot 1-story brick building.

A view of the 60-cycle side of the building is shown in Figure 10. The 60-cycle transformer of unit *2* and one of the d-c reactors are at the left. Figure 11 is a view of the building interior, showing the equipment of unit *2*. The 60-cycle rectifiers are at the left of the aisle, the 25-cycle rectifiers are at the right. The control cubicles are placed behind the fence between the rectifiers. The cores of the anode-paralleling reactors and anode-current transformers are mounted above the rectifiers. The d-c breakers are located on balconies. The control switchboards and the operator's desk are at the opposite end of the building, adjacent to unit *1*.

The station also has a conditioning stand with three spare excitron tubes, the necessary auxiliaries, degassing equipment, and facilities for reconditioning tubes. These facilities are utilized also for reconditioning and degassing excitron tubes for 250-volt rectifier units installed in other parts of the plant.

EFFICIENCY AND POWER FACTOR

The efficiency curve for unit *2* of the Gary frequency-changer installation is given in Figure 12. It is the calculated over-all efficiency between the 25-cycle and 60-cycle line terminals, including the losses of the rectifier auxiliaries. The curve is fairly flat over the greater part of the load range, which is typical for rectifier units.

The calculated power-factor curves on the 60-cycle side of a rectifier unit, when operating as a power rectifier and a power inverter, are given in Figure 13. The power factor is the ratio of kilowatts to kilovolt-amperes, which includes the effect of the harmonic components in the current. As was explained in another part of the paper, the output current of the power inverter has a leading power factor, while the input to the power rectifier has a lagging power factor. The increase in the power factor of the inverter at increasing load is due to an increase in the commutating angle. Referring to Figure 3, as the commutating angle increases with an increase in the current, the center of the anode current wave shifts to the right, and the angle ϕ decreases.

The power factor of the power rectifier is lowered when its output voltage is reduced by phase control, in approximately the ratio of the voltages. Its power factor is therefore variable, depending on the amount of phase control required for regulating the load of the frequency changer.

OPERATION

The Gary frequency changer has been operated mostly for transmitting power from the 25-cycle to the 60-cycle system. At times it operates in the opposite direction for extended periods, when some of the 25-cycle generating equipment is shut down for maintenance or for other reasons. On some days the direction is alternated.

The units are operated generally on base-load control, and the load level is set by the operator, at the direction of the steel company's power dispatcher, to meet the varying power requirements. Sections of the voltage and kilowatt graphic charts for one of the units are shown in Figure 14. The recording wattmeters have a zero-center scale, for recording either direction of power flow. The kilowatt records for the two sides of the frequency changer are therefore a reflection of each other, and the small difference in their magnitudes is due to the losses in the unit.

The transmitted power is kept steady, at the levels set by the operator, through the action of the load regulators, in spite of the voltage fluctuations on both systems. The rapid dips in the 60-cycle voltage were caused by the load of a large rolling mill. A "pip" in the kilowatt chart near 2 a.m. was produced by a suppressed arc-through on the power inverter.

Since the excitron rectifiers were installed, unit *1* of the frequency changer has been in operation a little over 4 years, and unit *2* over 3 years. During that time 12.5 per cent of the excitron tubes were reconditioned. The thyratron tubes used for firing the rectifier grids have been the major maintenance item, because of their relatively short useful life, averaging 4 to 5 months. It is expected that improvements in the tubes will lengthen their life considerably. They are operated at a fraction of their rated current.

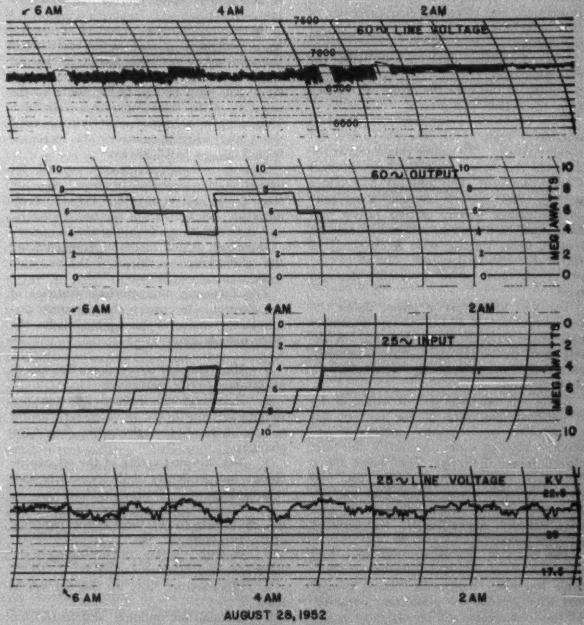


Figure 14. Graphic charts of line alternating voltages and kilowatts for one 8,000-kw unit of Gary frequency-changer installation

There have been occasional trip-outs of a unit caused by arc-backs, but it is restarted immediately, and the outage time is usually less than a minute.

Inductive co-ordination tests were conducted in 1945 to study the influence of the frequency changer on the power systems from the viewpoint of telephone interference. The voltage telephone influence factor (TIF) on the 60-cycle transmission and distribution system of the Northern Indiana Public Company was not appreciably affected by the operation of the frequency changer.

SUMMARY

The electronic frequency changer has proved itself as a practical and highly flexible tie between a-c power systems. The interchange of power can be controlled easily and accurately, unaffected by frequency and voltage variations. However, its application opportunities are limited. Because of its higher first cost, it can be applied economically only where a synchronous-synchronous frequency changer cannot be used,[7] as was the case for the Gary installation.

Besides the Gary installation, the United States Steel Corporaton has an electronic frequency changer operating in a steel mill near Pittsburgh. It is rated at 20,000 kw and employs a different type of rectifier tube.[8] It was also installed in 1943. The engineers of the steel company are to be commended for pioneering these applications.

References

1. HOW INVERTER TRANSFERS D-C POWER, **R. W. Holman.** *Electrical World* (New York, N. Y.), volume 125, number 15, April 13, 1946 pages 112–16.

2. AMERICAN STANDARD FOR POOL CATHODE MERCURY-ARC POWER CONVERTERS. *Publication Number C34.1—1949*, American Standards Association (New York, N. Y.), 1949.

3. MERCURY ARC POWER RECTIFIERS (book), **O. K. Marti, H. Winograd.** McGraw-Hill Book Company, Inc., New York, N. Y., 1930, page 127.

4. ELECTRICAL ENGINEERS' HANDBOOK (book), **H. Pender, W. A. Del Mar.** John Wiley and Sons, Inc., New York, N. Y., fourth edition, POWER RECTIFIERS AND INVERTERS, **H. Winograd,** section 11, 1949, page 27.

5. A FAST CIRCUIT BREAKER, **D. I. Bohn, Otto Jenson.** *AIEE Transactions* (*Electrical Engineering*), volume 61, March 1942, pages 165–68.

6. DEVELOPMENT OF EXCITRON-TYPE RECTIFIER, **H. Winograd.** *AIEE Transactions*, volume 63, 1944, pages 969–78.

7. FREQUENCY CHANGERS—CHARACTERISTICS APPLICATIONS, AND ECONOMICS, **S. B. Crary, R. M. Easley.** *AIEE Transactions* (*Electrical Engineering*), volume 64, June 1945, pages 351–58.

8. THE ELECTRONIC CONVERTER FOR EXCHANGE OF POWER, **F. W. Cramer, L. W. Morton, A. G. Darling.** *AIEE Transactions*, volume 63, 1944, pages 1059–69.

◆

Discussion

W. E. Gutzwiller (Allis-Chalmers Manufacturing Company, Milwaukee, Wis.): Mr. Winograd covered, in a very thorough manner, the principle of operation, design, protection, and electrical characteristics of this 16,000-kw electronic frequency-changer installation.

This discussion deals briefly with the economic aspects of this installation as they prevailed at the time the order for the original equipment was placed.

This electronic frequency changer was purchased mainly for 60-cycle power demand control. The power generated in the 25-cycle system by the United States Steel Corporation was then low-cost by-product power. The 60-cycle system, using purchased utility power, carried heavy mill loads, such as hot-strip, cold-reduction, and temper mills, causing wide fluctuations in the 60-cycle power demand. The purchased power was metered on a $^1/_2$-hour demand basis.

For best utilization of power under the power contract, it was necessary to control closely the 60-cycle power demand. Without such control the mills could frequently exceed the normal controlled demand by 15 to 20 per cent.

There was a definite need to utilize in the 60-cycle system some of the low-cost 25-cycle power produced by unstorable by-product fuel, in order to control the purchased power demand. This called for a frequency changer tie between the two power systems, whose input into the 60-cycle system could be controlled quickly and automatically, depending on the metered 60-cycle power.

The 25-cycle system at that time had ±1.5-cycle, and the 60-cycle system ±0.1-cycle frequency variations. Each system had over 200,000-kw generating capacity. All these conditions still prevail more or less today.

A frequency changer of the nonsynchronous tie type only could be considered. It was then estimated that for the initial installation a set of approximately 7,000 kw would show the maximum economy for demand control. The set also had to provide for operation, in emergency, from 60 to 25 cycles, to take care of temporary shortages of generating capacity in the 25-cycle system.

After careful consideration of all aspects by the engineers of the United States Steel Corporation, an electronic frequency changer was chosen. The original set was rated 6,667 kw, with provision for doubling the capacity of the installation at a later date. This expansion was carried out about 5 years later, when the capacity was increased from 6,667 kw to 16,000 kw.

The system requirements essential for operation of an electronic frequency changer, that is, the a-c voltage for inverter commutation, the kilovars of the power system, and the excitation kilovars of the frequency changer set, which must be supplied from the synchronous machines of the two systems, were met at Gary.

The following characteristic features of the electronic frequency changer were factors in the choice of this equipment for the Gary installation:

1. High over-all conversion efficiency is maintained over a wide range of load.
2. Relatively wide frequency fluctuations in both systems do not affect power transfer or stability of the electronic frequency changer.
3. The kilowatt rating of the set can be chosen independently of the respective system capacities.
4. The electronic frequency changer does not contribute to the short-circuit capacity of the two systems.
5. Starting, loading, and unloading of the set is easy and relatively quick.
6. Maintenance cost of the station is low.
7. Installation, building, and ventilation costs are low because of relatively light weights, absence of heavy rotating masses, and water cooling.
8. Total installed cost of the electronic set was favorable in comparison with similarly rated nonsynchronous-type rotating frequency-changer sets.

The operating records of this installation as of today indicate that most, if not all of the original expectations were realized.

P. A. Jeanne (Bell Telephone Laboratories, Inc., New York, N. Y.): The installation described in Mr. Winograd's paper and the one in reference 8 of the paper afforded an excellent opportunity to study the inductive co-ordination aspects of electronic frequency changers. Extensive observations of inductive influence were made by representatives of the Joint Subcommittee on Development and Research of the Edison Electric Institute and Bell System with the co-operation of the manufacturers and users of the equipment and the interested utilities. As stated in the paper under discussion, the frequency changer did not appreciably affect the TIF of the 60-cycle transmission system supplying it. A review of the test data indicates that there was a very substantial increase in inductive influence on the 6.9-kv bus directly supplying the frequency changer. However, the distribution of harmonics among the various transmission circuits forming the power network supplying the 6.9-kv bus resulted in only a small effect on the influence of the transmission circuits. If there had been any extensive telephone-circuit exposures to lines fed directly from the 6.9-kv bus, noise problems may have been encountered. The inductive influence for a given load condition was somewhat higher on the power system when it was receiving power from the frequency changer than on the same system when it was supplying power to the frequency changer. This is attributable to the greater amount of grid control used with the inverter, which causes more distortion than when the same unit is operating as a rectifier unit.

These observations are based on tests made at the time of the original installations.

Doubling of the frequency changer capacity of the Gary installation might be expected to increase the inductive influence. However, the second unit was so connected that when both are operating, the equivalent of a 24-phase rectifier in place of a 12-phase rectifier is obtained on the 60-cycle side, and a 12-phase rectifier in place of a 6-phase rectifier on the 25-cycle side. This is quite beneficial in reducing the inductive influence because it reduces by one-half the number of theoretically unsuppressed harmonics and they are of higher harmonic order, where their magnitudes are less.

Both this installation and that of reference 8 of the paper are excellent examples of the employment of the multiphasing principle to reduce wave-shape distortion and thus the inductive influence of rectifiers. We in the communication field greatly appreciate the attention given to this matter by both manufacturers and users of large rectifier installations.

I. K. Dortort (I–T–E Circuit Breaker Company, Philadelphia, Pa.): The author has covered the subject very completely and there is nothing I can add except to point out a few pitfalls.

The voltage charts of the two a-c systems shown in Figure 14 of the paper are typical, but more violent fluctuations occur than can be seen on these charts. If, in Figure 4 of the paper, a line parallel to E_{d1}, indicating the output voltage of the rectifier at maximum a-c voltage and no control and another line representing minimum a-c voltage and maximum phase control were drawn, and then two similar lines parallel to the inverter voltage were drawn, there would be formed a parallelogram including the area of operation possible when both system voltages vary violently between the limits indicated. It must be remembered that no stable regulator, even of the electronic type, can, in practice, be depended on to correct the power transfer for very sudden changes in voltage since corrective action must be taken within a small fraction of a cycle. The d-c reactors can be made large enough, at considerable expense, to slow up the rate of change and give the regulator a chance to catch up. Taking into account the speed of the regulator and the flywheel effect of the d-c reactor, the parallelogram defining the limits of operation can be constricted. For safe operation, however, the right-hand corner of the parallelogram must fall within the safe-load limit of the inverter and the left-hand corner must be greater than the minimum load limit imposed by the stability of the inverter, the system, or particular relaying schemes.

Although not mentioned in this paper, static condensers can be used to supply the commutating voltage, making it possible for the frequency changer to supply an independent load. While this method is possible, difficulties in starting and lack of stability under widely varying operating conditions make the system impractical for large installations.

Power factor of the inverter has been clearly defined in several different ways throughout the paper and yet, it seems to me, more emphasis is needed to correct erroneous thinking on this matter.

It is demonstrated that the inverter must operate at leading power factor in order to commutate successfully; that is to say, the connected system must accept leading current. It is clear that a leading power-factor load is, at the same time, a source of lagging reactive power. In other words, lagging kilovars must be supplied to the inverter by connected generators, or other rotating equipment capable of performing this function, or by static condensers. An electronic frequency changer does not improve the power factor of a system.

E. W. Brewer (General Electric Company, Pittsburgh, Pa.): Mr. Winograd's excellent paper is particularly interesting to many in the Pittsburgh district. A similar installation, which was mentioned in the paper, has been operating for 10 years at the Edgar Thomson Works of the United States Steel Corporation at Braddock, Pa. A few comments on the field experience with this unit may be of interest, as it confirms the practicality of this type of equipment.

The system conditions are similar to those of Gary in that power is generated at both 25 cycles and 60 cycles, supplemented with purchased power at 60 cycles.

The electronic, reversible, nonsynchronous tie consists of two units having a combined rating of 20,000 kw continuously, with overload ratings up to 40,000 kw for 1 minute. A-c power from the 60-cycle 66-kv system is converted by the rectifier to d-c power at approximately 15,000 volts, and then inverted to a-c power in the 25-cycle 44-kv system.

The amount and direction of power flow is remotely controlled. The usual system requirements are such that about 10,000 to 15,000 kw is supplied through the frequency changer to the 25-cycle side. At times as high as 24,000 kw is supplied for approximately 1 hour, and on one occasion, over 30,000 kw was supplied for 1 hour, with a close watch being maintained on tube temperatures. It is operated on standby at loads as low as 500 kw per unit, and during emergencies, 60-cycle power is supplied. Commutating reactors were used originally, but were eliminated when it was found they were not necessary.

Each 10,000-kw unit contains 24 type *GL-506* sealed-pentode ignitron-type tubes. It was expected originally that the tube life might be only 10,000 hours. The latest survey (in January 1953) disclosed, however, that 16 of the original tubes were still in service. The average age (based on actual operating time) of the 48 tubes now installed in the two units was 5.6 years. The average age of the tubes being used has been increasing and, as more of the original tubes reach the end of their life, accurate values of average tube life can be calculated. These values will exceed the average age of tubes in service by a substantial margin.

The river water side of the heat exchangers is cleaned once or twice a year. As with tube changes, this operation is performed on one unit while the other unit is operating.

Harold Winograd: Mr. Gutzwiller has supplied additional background information on the economic factors of the Gary frequency changer installation. Reference 1 of the paper contains interesting data on the application of the frequency changer for controlling the maximum demand.

Mr. Jeanne has pointed out correctly that there was a substantial increase in the voltage TIF at the 6.9-kv 60-cycle terminals of the frequency changer, which was due to the reactance of the 132/6.9-kv step-down transformer bank. This was of no consequence in the Gary installation because the 6.9-kv circuit was inside the plant and no exposures of telephone circuits were involved. From the viewpoint of inductive co-ordination, the significant fact is that an electronic frequency changer has the same effect and can be treated in the same way as any other rectifier installation of comparable size and operating with a corresponding amount of phase control. Where an inductive co-ordination problem is expected, corrective measures, such as phase multiplication, can be applied in the same manner as for rectifier installations in electrochemical service.

The problem of load control due to variations of the a-c line voltages is nowhere near as critical as indicated by Mr. Dortort. It is neither necessary nor desirable to set the angle of advance of the power inverter for extreme and infrequent swings of the a-c voltages. Such a setting would have an adverse effect on the power factors and on the performance of the frequency changer. The load regulator or the protection system can take care of these conditions.

The d-c reactors, of the size used in the Gary installation, have a decided stabilizing effect, and the regulator does not have to take corrective action "within a small fraction of a cycle." For example, with a 10-millihenry reactor in the d-c circuit, a sudden 20-per-cent dip in the inverter voltage will produce an initial rate of increase in the direct current of about 25 per cent of rated current, per cycle (on a 60-cycle basis). Assuming an angle of advance of 35 degrees and a transformer commutating reactance of 6 per cent, the commutation-limit load during such a voltage dip would be reached in about 4 cycles, allowing a margin angle of 15 degrees for deionization of the tube grids. Such a large voltage dip on either system, or a combined simultaneous voltage change in opposite directions on the two systems, would be unusual, but there is sufficient time for a reasonably good regulator to take corrective action.

Electronic frequency changers of relatively low rating, in which the commutating voltage is generated by capacitors in combination with an inductive load, have been applied for conversion of 60-cycle power to about 1,000 cycles for induction heating. As stated by Mr. Dortort, this type of frequency changer would not be very stable under fluctuating load conditions.

Regarding the power factor and kilovars, it has been explained in the paper that the output current of the power inverter has a leading power factor and that the a-c system "must be capable of supplying the lagging kilovars required by the load and the power inverter." Putting out leading kilovars is the same as receiving lagging kilovars.

Mr. Brewer has given some interesting information on the operation of the electronic frequency changer in the Edgar Thomson Works, which was installed in the same year as the first unit at Gary. Besides using different types of rectifier tubes, the two installations have several notable differences in their power circuits.

The rectifier transformers of the Edgar Thomson units have a considerably higher reactance, provided by series a-c reactors. This increased the variation of the commutating angle with load and made it necessary to control the angle of advance of the power inverter in function of the load. The higher reactance also tends to lower the power factor on the input and output sides.

In the Gary installation two rectifier tubes are connected in parallel to each phase of a rectifier transformer, while the Edgar Thomson units have one tube per phase. By paralleling two tubes, a misfire or excitation interruption on one tube of the power inverter is not likely to cause an arc-through because the parallel tube continues to commutate the current, even though it has to conduct the phase current alone during the misfire interval.

The d-c reactors of the Gary units tend to limit the fault current during arc-throughs and aid in suppressing them. The switching equipment in the d-c circuits, independent cooling systems for the individual rectifiers, and individual current regulators for each section permit operating one section of a Gary unit while the other is out of service. This provides greater operating flexibility.

The use of a double-way rectifier circuit for the Edgar Thomson units made it possible to employ simpler rectifier transformers, although some complication was added to the installation because the tubes of each rectifier had to be insulated from each other, and insulating transformers had to be used for the excitation and grid-control circuits of the individual tubes.

This is a rather brief and possibly 1-sided comparison of a few salient technical differences in the two installations. It is interesting that successful solutions of an engineering problem were attained from different approaches.

Oscillatory Circuits Containing Iron-Cored Inductances

ABD EL-SAMIE MOSTAFA
NONMEMBER AIEE

M. EL-SHISHINI
NONMEMBER AIEE

Synopsis: The purpose of the present paper is the development of a theory for analyzing nonlinear iron-cored inductance oscillatory circuits where the forced behavior of the system is considered.

The methods of superposition and successive approximation are applied to the solution of circuits having periodically varying capacitances and the results are directly applied to the present nonlinear problem. The solution obtained gives the harmonics present, a matter which has not been considered by previous authors.

A complete analysis of the properties of oscillatory circuits containing iron-cored inductances is given. Performance and design of static stabilizers using such circuits are also given. Experimental verification has been carried out, and the results agree well with the theoretical predictions.

ELECTRIC circuits containing iron-cored inductances have become of increasing importance due to their extension into many fields of electrical engineering. The property of sudden variation, the so-called critical points, makes ferroresonant circuits useful as sensitive elements in relay and control apparatus.[1–7] Magnetic generation of a group of harmonics,[8] for supplying carrier currents to various multichannel carrier telephone systems and for frequency comparison and standardization, employ such circuits. Static stabilizers are good examples of these circuits. Therfore, a need is felt for a rigorous but simplified treatment for quantitative evaluation of oscillatory circuits containing iron-cored inductances.

Several papers for the explanation and predetermination of the oscillations in these circuits have been written.[1–8] In all of these papers, the nonlinear characteristic of the iron-cored inductance was treated from the point of view of effective values and not from the point of view of instantaneous values. Therefore, previous authors' analysis is similar to the secant method used in the determination of the amplitude of oscillation in a simple electron tube oscillator and which was proved to be unsatisfactory.[9] Moreover, previous writers ignore the harmonics which may be of considerable amplitudes.

In the present analysis, the nonlinear inductance characteristic is treated from the point of view of instantaneous values and the solution obtained contains the harmonics as well.

Analysis of Iron-Cored Reactor Oscillatory Circuits

The circuit to be analyzed is the ferroresonant parallel circuit, see Figure 1. In most of the iron-cored reactors used in this connection, both the hysteresis and eddy-current losses can be approximately represented by a constant resistance R' shunting the inductor, so long as the frequency of operation is low.

According to Kirchhoff's first law

$$\sum i=0 \text{ at point } P$$

therefore

$$i_C+i_{R'}+i_L=i$$

giving

$$C\,dv/dt+v/R_{\text{tot}}+\left(\int v\,dt\right)/L_t =(V_s/R)\sin(\omega t+\theta) \quad (1)$$

where, i_C = current flowing through condenser $= C\,dv/dt$; v = voltage across coil $\cong d\phi/dt=d(L_ti_L)/dt$, neglecting winding resistance of coil; ϕ = coil flux linkage; L_t = instantaneous inductance $=\phi/i_L$; $i_L=(\int v\,dt)/L_t$ = current flowing through the iron-cored reactor; i = current passing through load resistance $R=(v_s-v)/R$; v_s = supply voltage $=V_s\sin(\omega t+\theta)$; $i_{R'}=v/R'$ = current corresponding to iron loss; $1/R_{\text{tot}}=(1/R)+(1/R')$.

Figure 2 shows instantaneous (that is, d-c) $i\rightarrow\phi$ characteristic passing through the tips of the magnetizing cycles as well as the $(1/L_t)\rightarrow\phi$ characteristic. The first portion of the $i\rightarrow\phi$ characteristic can be satisfactorily considered linear as shown by the dotted line. The corresponding portion in the $(1/L_t)\rightarrow\phi$ characteristic is also shown dotted. $(1/L_t)\rightarrow\phi$ characteristic can be represented by either a power series or a trigonometric series. However, Fourier representation is general and can be used whether the representation is a function of ϕ or i. In the latter case, the power series is very tedious as it requires a large number of terms for satisfactory representation. The characteristic is symmetrical about the vertical axis, and therefore only cosine terms will appear in the Fourier analysis, giving

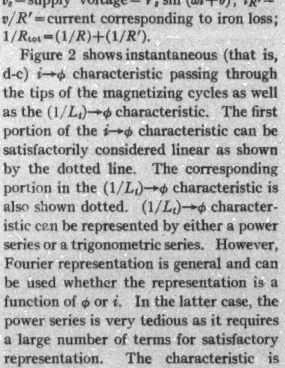


Figure 1. Ferroresonant circuit to be analyzed

Paper 53-156, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted April 3, 1951; made available for printing December 4, 1952.

ABD EL-SAMIE MOSTAFA is with the Alexandria University, Alexandria, Egypt. M. EL-SHISHINI is a member of the Egyptian Senate, Cairo, Egypt.

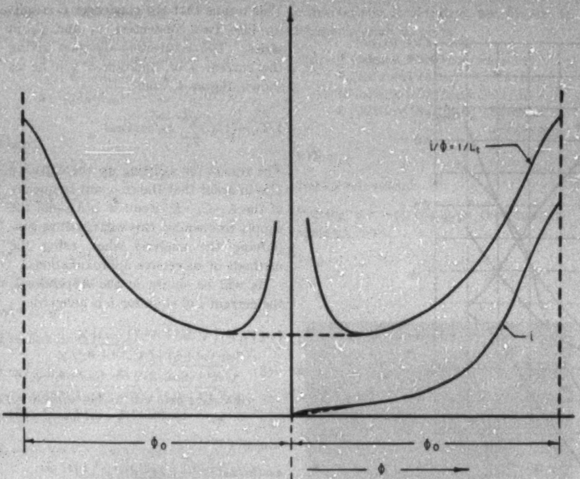


Figure 2. Current as well as $(1/L_t)$ versus flux linkage

$$1/L_t = a_0 + \sum_{n=1} a_n \cos(n\pi\phi/\phi_0)$$

where a_0, a_1, a_2, ... a_n are constants to be determined from Fourier analysis. The upper limit of the summation depends on the number of terms in the expansion required for satisfactory representation, but generally it is infinity.

From the nature of the problem, the steady-state flux must contain only odd harmonics. Let the steady-state flux linkage ϕ be given by

$$\phi = -\phi_m \cos\omega t + \phi_3 \cos(3\omega t + \theta_3) + \phi_5 \cos(5\omega t + \theta_5) + \ldots$$

giving

$$v = d\phi/dt = \phi_m\omega \sin\omega t - \phi_3 3\omega \sin(3\omega t + \theta_3) - \phi_5 5\omega \sin(5\omega t + \theta_5) + \ldots$$

The angle θ will consequently be the phase angle between the supply voltage v_s and the fundamental of v at the steady state. The negative sign in ϕ is simply used to give a positive sign, for the fundamental of v, on differentiation.

For the solution of equation 1 to the first power of the parameters a_n of the nonlinear inductance, it is sufficient to substitute $-\phi_m \cos\omega t$ for ϕ in the Fourier analysis. This approximation is sufficient for most practical cases. The resulting solution will contain the harmonics to a satisfactory degree of approximation. Substitution gives

$$1/L_t = a_0 + \sum_{n=1} a_n \cos(\phi_n \cos\omega t) = (1/L_0)(1 + K_2 \cos 2\omega t + K_4 \cos 4\omega t + \ldots) \quad (2)$$

where

$$1/L_0 = a_0 + \sum_{n=1} a_n J_0(\phi_n)$$

$$\phi_n = n\pi\phi_m/\phi_0$$

$$K_2 = \left\{-2\sum_{n=1} a_n J_2(\phi_n)\right\} / (1/L_0)$$

$$K_4 = \left\{2\sum_{n=1} a_n J_4(\phi_n)\right\} / (1/L_0) \ldots$$

J_0, J_2, J_4 ... are Bessel functions of zero, 2nd, 4th, ... orders respectively.[10]

In most of the iron-cored reactors used, equation 2 is highly convergent in the normal operating ranges and usually terms up to K_4 are sufficient for satisfactory results. Substituting equation 2 in 1 gives

$$Cdv/dt + v/R_{tot} + (1/L_0)(1 + K_2 \cos 2\omega t + K_4 \cos 4\omega t + \ldots)\int v dt = (V_s/R) \times \sin(\omega t + \theta) \quad (3)$$

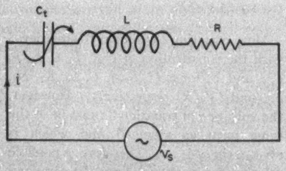


Figure 3. Series oscillatory circuit having a periodically varying capacitance

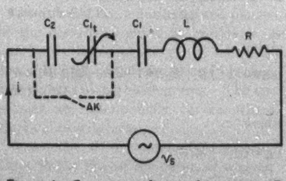


Figure 4. Connection diagram for equation 5

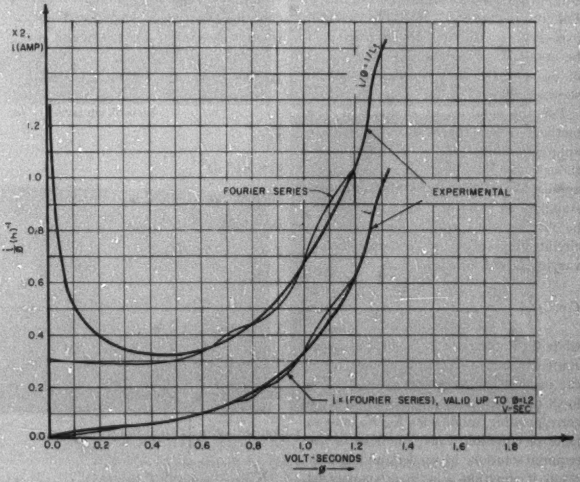


Figure 5. Fourier representation of $i-\phi$ as well as $(1/\phi)-\phi$ characteristics

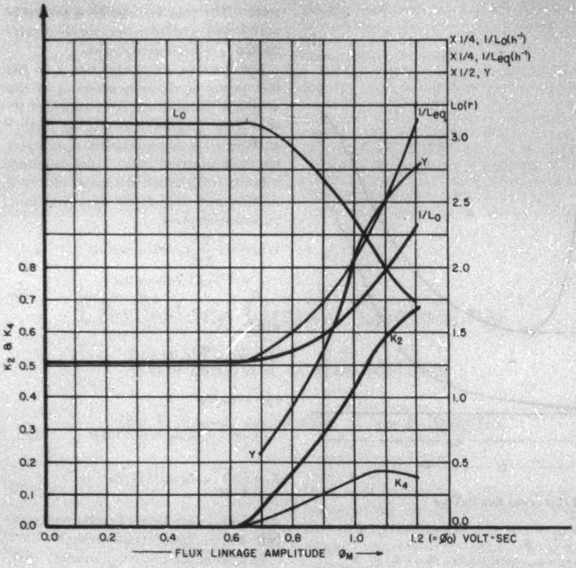


Figure 6. Core characteristics

In this way, the nonlinear differential equation 1 has been transformed to a linear one with periodically varying coefficients. This linearization concerns only the differential equation. On the other hand, the solution of equation 3 will not give the amplitude ϕ_m explicitly. For given values of V_s, R, C, and a certain iron-cored inductance, the solution can be obtained by assuming several values of ϕ_m and solving[11] for V_s, as will be shown later.

For the solution of equation 3, the methods of superposition and successive approximation will be adopted. Although these methods can be directly applied to equation 3, yet a better understanding of the nature of the solution can be obtained when solving the dual series circuit, Figure 3, where C_t is a periodically varying capacitance and given by

$$C_t=C/\left\{1+\sum_{n=1}^{\infty} h_{2n}\cos 2n\omega t\right\}$$

while the inductance L and the resistance R are constants. Here the amplitudes of the capacitance variation h_2, h_4, and so forth of Figure 3 are constants while the corresponding amplitudes K_2, K_4, and so forth of equation 3 are functions of the required solution. Misunderstanding may result if equation 3 is directly analyzed using the methods of successive approximation. The series forming C_t of Figure 3 is assumed convergent, which is the case of equation 3.

The circuit is supplied with a voltage v_s given by $V_s\sin(\omega t+\theta)$ and $(\theta-\pi/2)$ is the phase angle between the applied voltage v_s and the fundamental of the capacity variation. The differential equation for the current i in the circuit is

$$Ldi/dt+Ri+(1/C)\left\{1+\sum_{n=1}^{\infty} h_{2n}\cos 2n\omega t\right\}\times \int idt=V_s\sin(\omega t+\theta) \quad (4)$$

Let

$$\omega_0=\omega(1+d)$$

where

$$\omega_0^2=1/LC$$

d = amount of detune

giving

$$\omega_0^2=\omega^2(1+d)^2=(1/LC_1)+(1/LC_2)$$

$$1/C_1=\omega^2L,$$

$$1/C_2=(2d+d^2)/C_1=\omega^2L(2d+d^2)$$

Substitution in equation 4 gives

$$Ldi/dt+Ri+\left\{(1/C_1)+(1/C_2)+(1/C)\times \sum_{n=1}^{\infty} h_{2n}\cos 2n\omega t\right\}\int idt=V_s\sin(\omega t+\theta) \quad (5)$$

This means that the condenser C is split up into two condensers C_1 and C_2 in series. The connection diagram giving the current i of equation 5 will be as shown, Figure 4, where

$$1/C_{1_t}=(1/C)\sum_{n=1}^{\infty} h_{2n}\cos 2n\omega t$$

The reason for splitting up the capacity C is in order that the resonant frequency of the $L-C_1-R$ circuit would equal the supply frequency; this will result in simplifying the analysis when using the methods of successive approximations.

As will be shown in the Appendix I, the current i of equation 5 is given by

$$\begin{aligned}i=(V_s/R)\{\sqrt{M^2+N^2}/(1-a)\}\times \\ [\sin(\omega t+\theta_1)+(\sqrt{A_3^2+B_3^2}/\times \\ \sqrt{R^2+(8\omega L/3)^2})\cos(3\omega t+\theta_3-\phi_3)+ \\ (\sqrt{A_5^2+B_5^2}/\sqrt{R^2+(24\omega L/5)^2})\times \\ \cos(5\omega t+\theta_5-\phi_5)+\ldots] \quad (6)\end{aligned}$$

where

$$\begin{aligned}a&=(h_2/2\omega CR)^2-\{(2d+d^2)/\omega C_1R\}^2 \\ &=(\omega L/R)^2[\{(h_2/2)(1+d)^2\}^2-(2d+d^2)^2]\end{aligned}$$

$$M=\cos\theta+(\omega L/R)\{(h_2/2)(1+d)^2-(2d+d^2)\}\sin\theta$$

$$N=\sin\theta+(\omega L/R)\{(h_2/2)(1+d)^2+(2d+d^2)\}\cos\theta$$

$$\tan\theta_1=N/M$$

$$A_3=\{(h_2+h_4)/2\omega C\}\cos\theta_1$$

$$B_3=\{(h_2-h_4)/2\omega C\}\sin\theta_1$$

$$\tan\theta_3=B_3/A_3$$

$$\tan\phi_3=8\omega L/3R$$

$$A_5=(h_4/2\omega C)\cos\theta_1$$

$$B_5=(h_4/2\omega C)\sin\theta_1$$

$$\tan\theta_5=B_5/A_5=\tan\theta_1$$

$$\tan\phi_5=24\omega L/5R$$

Now, equation 4 is similar to equation 3 with the substitution of i, R, L, C, h_2, h_4 . . ., and V_s for v, $(1/R_{tot})$, C, L_0, K_2, K_4, . . . and (V_s/R) respectively. However, the voltage v of equation 3 must be of the same form as assumed and which is $v=\phi_m\omega\sin\omega t+\ldots$ Therefore, a possible solution for equation 3 can be obtained from equation 6 provided $\theta_1=0$, that is

$$N=0=\sin\theta+(\omega L/R)\{(h_2/2)(1+d)^2+(2d+d^2)\}\cos\theta$$

giving

$$\tan\theta=(\omega L/R)\{(h_2/2)(1+d)^2+(2d+d^2)\}=\{\omega L-(1/\omega C_{eq})\}/R$$

where

$$C_{eq}=C/(1+h_2/2)$$

Therefore, the voltage v of equation 3 will be given satisfactorily by

$$v=V_s(R_{tot}/R)\cos\theta[\sin\omega t+\{(K_2+K_4)/\times 2\sqrt{(\omega L_0/R_{tot})^2+(8\omega^2/3\omega_0^2)^2}\}\times \cos(3\omega t-\phi_3)+\{(K_4/\times 2\sqrt{(\omega L_0/R_{tot})^2+(24\omega^2/5\omega_0^2)^2})\times \cos(5\omega t-\phi_5)+\ldots] \quad (7)$$

where

$\tan\theta=R_{tot}\{\omega C-(1/\omega L_{eq})\}$,

$R_{tot}=RR'/(R+R')$

$L_{eq}=L_0/(1+K_2/2)$

$\tan\phi_3=(8/3)\omega CR_{tot}$

$\tan\phi_5=(24/5)\omega CR_{tot}$

$\omega_0^2=1/L_0C$

The flux linkage

$$\phi=\int v\,dt \quad (8)$$

The voltage across the load resistance R will be given by

$$v_R=v_s-v=V_s\cos\theta[\sqrt{\tan^2\theta+(R_{tot}/R')^2}\times \sin(\omega t+\theta')-\{(R_{tot}/R)(K_2+K_4)/\times 2\sqrt{(\omega L_0/R_{tot})^2+(8\omega^2/3\omega_0^2)^2}\}\times \cos(3\omega t-\phi_3)+\{(R_{tot}/R)(K_4/2)/\sqrt{(\omega L_0/R_{tot})^2+(24\omega^2/5\omega_0^2)^2}\} \cos(5\omega t-\phi_5)+\ldots] \quad (9)$$

where

$\tan\theta'=(R'/R_{tot})\tan\theta=R'\{\omega C-1/\omega L_{eq}\}$

The load current is given by

$i=v_R/R$

As stated before, the solutions of equations 7, 8, and 9 are not direct. For given values of V_s, R, C and a certain iron-cored inductance, equation 8 can be solved by assuming several values of ϕ_m and solving for V_s. Interpolation or otherwise is normally required.

It should be noted that the predicted harmonics arrived at in equations 7, 8, and 9 are practically satisfactory estimates of the harmonic contents. For a closer approximation resubstitute $\{-\phi_m\cos\omega t+\phi_3\sin(3\omega t-\phi_3)+\phi_5\sin(5\omega t-\phi_5)+\ldots\}$ for ϕ in the Fourier analysis of $1/L_t$ where ϕ_3, ϕ_5 and so forth, ϕ_3, ϕ_5, and so forth are the amplitudes and phases of the harmonics given by equation 8; the resulting values of K_2, K_4, and so forth will now be modified, and solutions 7, 8, and 9 will still hold.

To test that the amplitudes and phases of the harmonics given by equations 7 and 9 are reliable estimates, the following examples will be calculated and checked experimentally by oscillograms. The examples are intentionally chosen where the nonlinearity is large.

From the $\phi\rightarrow i$ characteristic passing through the tips of the cycles of a given iron-cored inductance, the $(1/L_t)\rightarrow\phi$ characteristic is obtained, see Figure 5; harmonic analysis gives

$$1/L_t=0.4465-0.253665\cos(\pi\phi/\phi_0)+0.1481\cos(2\pi\phi/\phi_0)-0.0715\cos\times(3\pi\phi/\phi_0)+0.04275\cos(4\pi\phi/\phi_0)-0.03285\cos(5\pi\phi/\phi_0)+0.014665\times\cos(6\pi\phi/\phi_0)$$

where

$\phi_0=1.2$ volt-seconds

Putting $\phi=-\phi_m\cos\omega t$ in the harmonic analysis gives

$$1/L_t=(1/L_0)(1+K_2\cos 2\omega t+K_4\times\cos 4\omega t+\ldots)$$

$$1/L_0=0.4465-0.253665J_0(\pi\phi_m/\phi_0)+0.148085J_0(2\pi\phi_m/\phi_0)-0.0715\times J_0(3\pi\phi_m/\phi_0)+0.04275J_0(4\pi\phi_m/\phi_0)-0.03285J_0(5\pi\phi_m/\phi_0)+0.014665\times J_0(6\pi\phi_m/\phi_0)$$

$$K_2=2L_0\{-0.253665J_2(\pi\phi_m/\phi_0)+0.148085\times J_2(2\pi\phi_m/\phi_0)-0.0715J_2(3\pi\phi_m/\phi_0)+0.04275J_2(4\pi\phi_m/\phi_0)-0.03285\times J_2(5\pi\phi_m/\phi_0)+0.014665J_2(6\pi\phi_m/\phi_0)\}$$

$$K_4=2L_0\{-0.253665J_4(\pi\phi_m/\phi_0)+0.148085\times J_4(2\pi\phi_m/\phi_0)-0.0715J_4(3\pi\phi_m/\phi_0)+0.04275J_4(4\pi\phi_m/\phi_0)-0.03285\times J_4(5\pi\phi_m/\phi_0)+0.014665J_4(6\pi\phi_m/\phi_0)\}$$

The coefficients higher than K_4 were very

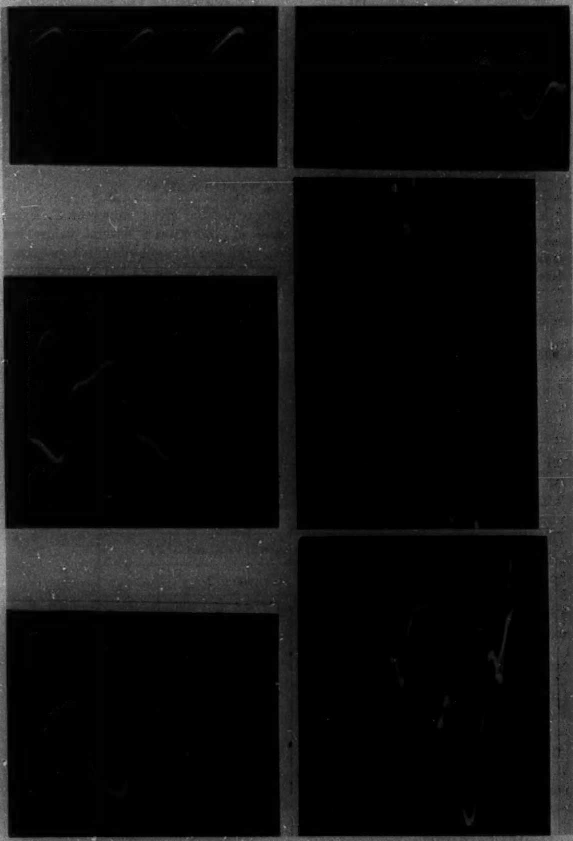


Figure 7. Experimental oscillograms of load voltage v_R and voltage v across reactor

Top: C=12 microfarads, V_s (rms)=264 volts, R=500 ohms
Middle: C=8 microfarads, V_s (rms)=300 volts, R=500 ohms
Bottom: C=10 microfarads, V_s (rms)=240 volts, R=500 ohms

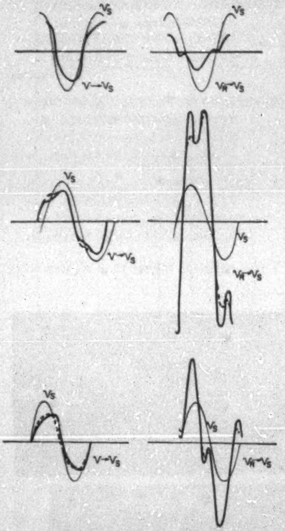


Figure 8. Wave forms of load voltage v_R and voltage v across reactor. Sinusoidal supply voltage is indicated. Theoretical wave forms are shown dashed

Top: C=12 microfarads, V_s (rms)=264 volts, R=500 ohms
Middle: C=8 microfarads, V_s (rms)=300 volts, R=500 ohms
Bottom: C=10 microfarads, V_s (rms)=240 volts, R=500 ohms

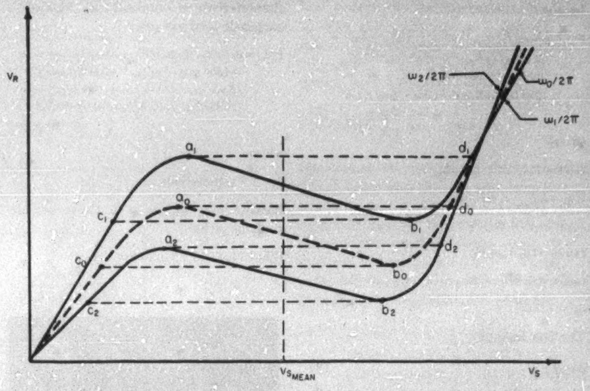


Figure 9. Load voltage versus supply voltage for three frequencies $\omega_0/2\pi$, $\omega_1/2\pi$, and $\omega_2/2\pi$ ($\omega_1>\omega_0>\omega_2$)

small and hence they were neglected. From the Bessel coefficients, the values of $1/L_0$, K_2, and K_4 can be obtained for any value of ϕ_m.

The characteristics of the core which are, $1/L_0$, K_2, K_4, L_0, $1/L_{eq}\{=(1+K_2/2)/L_0\}$ against ϕ_m are shown, see Figure 6.

The shunt resistance R' representing the iron loss of the present core=1,780 ohms.

EXAMPLE 1

Capacitance $C=12$ microfarads, resistance $R=500$ ohms, supply voltage $V_s=264$ volts rms, and supply frequency =40 cycles per second.

It is clear from equations 7 and 8 that the amplitude ϕ_m of the fundamental of the flux linkage is given by

$$\phi_m=(1/\omega)V_s(R_{tot}/R)\cos\theta$$

where

$$\tan\theta=R_{tot}\{\omega C-(1/\omega L_{eq})\}$$

$$R_{tot}=RR'/(R+R')$$

and

$$L_{eq}=L_0/(1+K_2/2)$$

substitute the given values of C, R, ω, and R' and assume several values for ϕ_m until one is found which is consistent; this value of $\phi_m=1.153$ volt-seconds. Using Figure 6, we get

$$L_0=1.836h$$

$$K_2=0.638$$

$$K_4=0.1716$$

$$1/L_{eq}=0.718h^{-1}$$

Substituting in equations 7 and 9 we get

$$v=290\{\sin\omega t+0.104\cos(3\omega t-72°)+0.0126\cos(5\omega t-80°)\}$$

$$v_R=84.8\{\sin(\omega t+15°)-0.356\cos\times(3\omega t-72°)-0.041\cos(5\omega t-80°)\}$$

$$v_s=373\sin(\omega t+3°20')$$

EXAMPLE 2

$C=8$ microfarads, $R=500$ ohms, $V_s=300$ volts rms, and supply frequency=40 cycles per second. Following the same lines as before we get

$$\phi_m=1.2 \text{ volt-seconds}$$

$$L_0=1.725h$$

$$K_2=0.675$$

$$K_4=0.153$$

$$1/L_{eq}=0.776h^{-1}$$

$$v=303\{\sin\omega t+0.164\cos(3\omega t-64°48')+0.018\cos(5\omega t-75°)\}$$

$$v_R=187\{\sin(\omega t-62°24')-0.2675\cos\times(3\omega t-64°48')-0.029\cos(5\omega t-75°)\}$$

$$v_s=424\sin(\omega t-23°)$$

EXAMPLE 3

$C=10$ microfarads, $R=500$ ohms, $V_s=240$ volts rms, and supply frequency=40 cycles per second, giving

$$v=264.8\{\sin\omega t+0.091\cos(3\omega t-69°)+0.013\cos(5\omega t-77°57')\}$$

$$v_R=77.7\{\sin(\omega t+16°40')-0.31\cos\times(3\omega t-69°)-0.044\cos(5\omega t-77°57')\}$$

$$v_s=340\sin(\omega t+3°48')$$

Figure 7 shows experimental oscillograms for v and v_R for the previous calculated cases. In the same figure the supply voltages are also shown, noting that they are 180 degrees out of phase in the v-oscillograms. The theoretical curves are drawn to the scale of the oscillograms and shown together for comparison in Figure 8; there is satisfactory agreement indicating that the amplitudes and phases of the harmonics given by Figures 7 and 8 are reliable estimates.

Some Properties of the Circuit

In order to simplify the analysis when getting the properties of the circuit, only the fundamental of the wave will be considered; it is clear from the previous examples that a discrepancy of the order of 10 per cent between experimental values (rms) and theoretical values (fundamental) is to be expected at the portions of the characteristics where the nonlinearity is high. However, most design data obtained in this way will be practically satisfactory. It should be noted here that considering only the fundamental and using the present analysis is totally different from the standard prac-

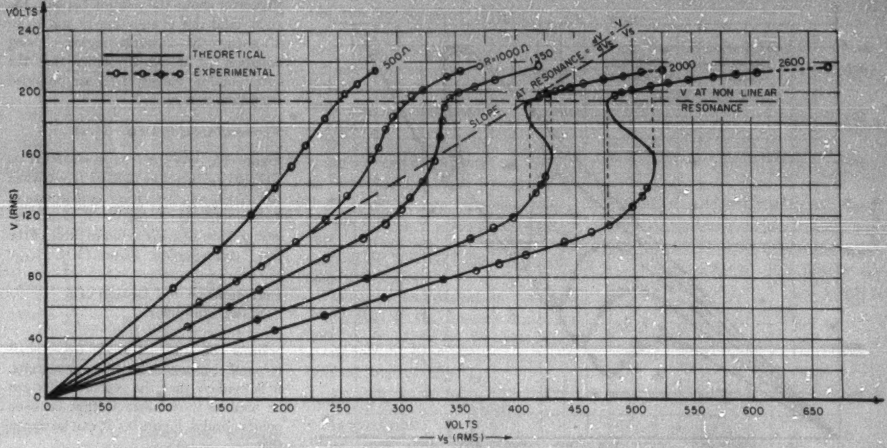


Figure 10. Characteristics of voltage V across reactor versus supply voltage V_s, C=10 microfarads

tice used by previous authors, namely, treating the nonlinear characteristic of the iron-cored inductance from the point of view of effective values.

The important properties of the circuit are the unstable regions and the point of nonlinear resonance.

The point of nonlinear resonance, is that at which v_s and the fundamentals of v and v_R are in phase, and hence $\theta=0=\theta'$, giving $\omega^2C=1/L_{eq}$. The corresponding value of ϕ_m is therefore independent of R, and can be obtained from $(1/L_{eq})\rightarrow\phi_m$ characteristic given in Figure 6. It can be shown easily that at the resonant point

$$V_s=(1+R/R')\omega\phi_m$$

$$V_R=(R/R')\omega\phi_m$$

$$dV_s/dV=1+R/R'$$

$$dR/dV=-(R+R')/V$$

where V is the amplitude of the fundamental of v. The unstable region is the region bounded by the locus $fV/dV_s=\infty$, that is, $dV_s/dV=0$. It can also be shown that

$$dV_s/dV=(1/\omega)dV_s/d\phi_m=(1+R/R')\times [1+R_{tot}^2\{(1/\omega L_{eq})-\omega C\}^2+ \phi_m R_{tot}^2\{(1/\omega L_{eq})-\omega C\}Y/\omega]/\times [1+R_{tot}^2\{(1/\omega L_{eq})-\omega C\}^2]$$

where $Y=d(1/L_{eq})/d\phi_m$ characteristic is given in Figure 6.

For

$$dV_s/dV=0$$

$$1/R_{tot}^2=-\{(1/\omega L_{eq})-\omega C\}\times \{(1/\omega L_{eq})-\omega C+\phi_m Y/\omega\} \quad (10)$$

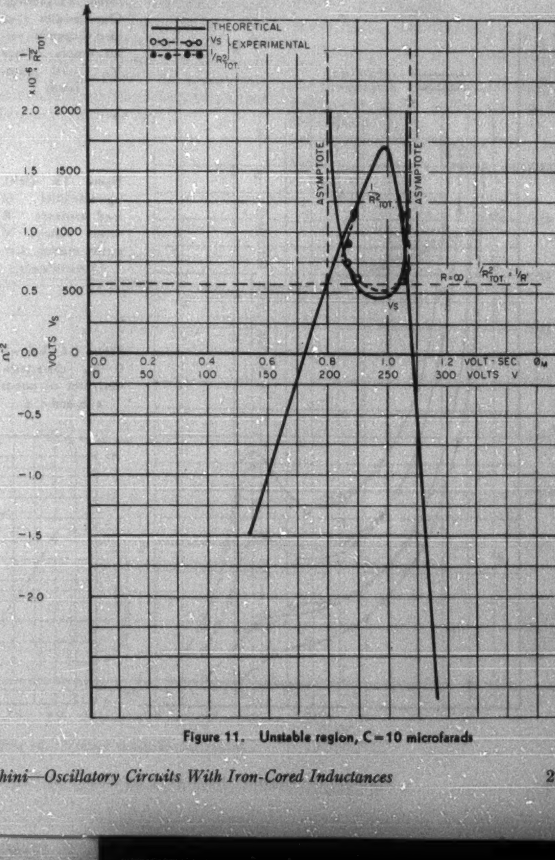


Figure 11. Unstable region, C=10 microfarads

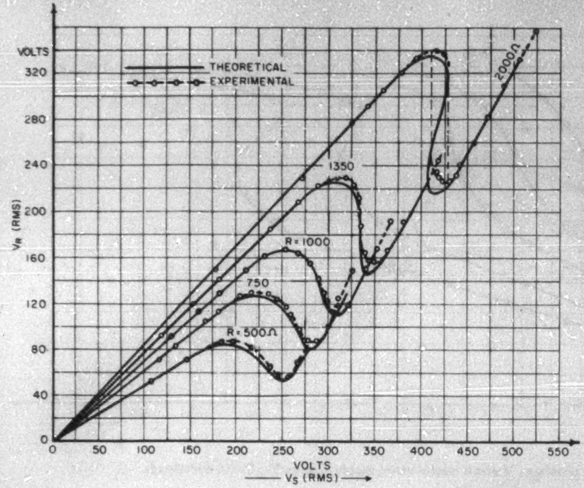


Figure 12 (above). Characteristics of load voltage V_R versus supply voltage V_s, C=10 microfarads

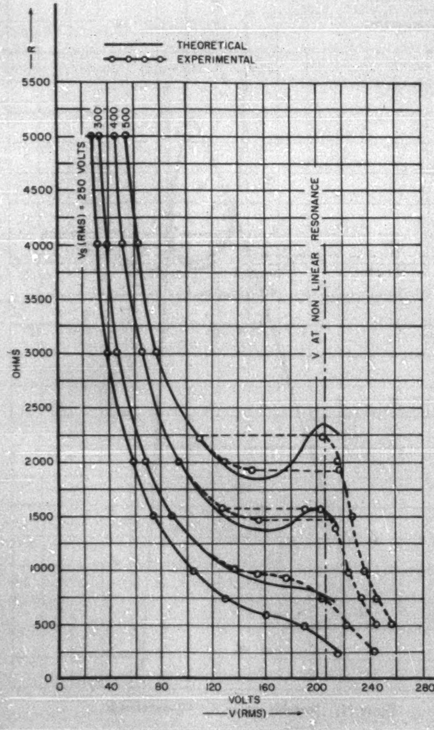


Figure 13 (left). Characteristics of load resistance R versus voltage V across reactor, C=12 microfarads

Figure 14 (below). C−ϕ_m characteristics, loci of points a, b, and c, d

Figure 11 shows the relation between the right-hand side of equation 10 and ϕ_m or V giving the unstable region. Using equation 10, it can be shown that the corresponding values of V_s are obtained from the relation

$$V_s=\omega\phi_m R\sqrt{\phi_m(Y/\omega)\{\omega C-(1/\omega L_{eq})\}},$$

Figure 11. As $(\partial R/\partial V)_{V_s}=(\partial R/\partial V_s)_V(\partial V_s/\partial V)_R$, therefore $(\partial R/\partial V)_{V_s}=0$ when $(\partial V_s/\partial V)_R=0$, indicating that points of sudden variation in the $V\rightarrow V_s$ characteristic, Figure 10, correspond to those of the $R\rightarrow V$ characteristic, Figure 13.

Performance and Design of a Static Stabilizer

From the study of the characteristics of parallel iron-cored oscillatory circuits, it is evident that the region which can be used as a stabilizer is that between points c and d, Figure 9. It can be shown that

$$V_R=\omega\phi_m R\sqrt{\{(1/\omega L_{eq})-\omega C\}^2+(1/R'^2)}$$

indicating that the line current I, given by $(I=V_R/R)$, depends only on ϕ_m and C, and is independent of R. It can also be shown that

$$V_s=\omega\phi_m \sqrt{\{(R+R')/R'\}^2+R^2\{(1/\omega L_{eq})-\omega C\}^2}$$

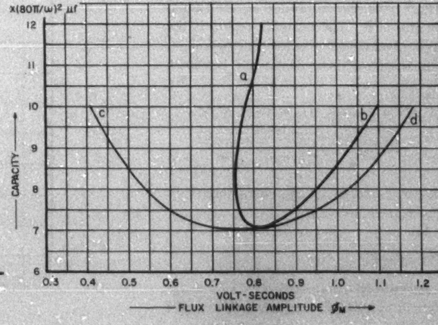
The points a and b are determined from the relation

$$dV_R/dV_s=0=(\partial V_R/\partial\phi_m)(\partial\phi_m/\partial V_s)$$

that is, from the relation, $dV_R/d\phi_m=0$, giving

$$\{(1/\omega L_{eq})-\omega C\}^2+(1/R'^2)+\{(1/\omega L_{eq})-\omega C\}Y\phi_m/\omega=0$$

showing that the flux linkages ϕ_{m_a} and ϕ_{m_b} at points a and b are independent of R. Again, as Y is positive as shown in Figure 6, $\{(1/\omega L_{eq})-\omega C\}$ must be negative in the region between a and b.



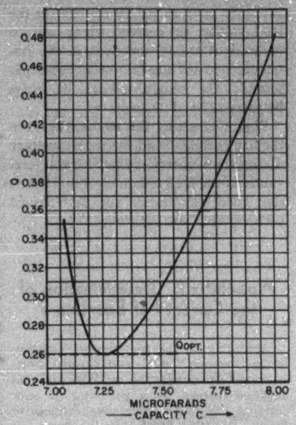


Figure 15. Q—C characteristic, optimum capacity for good stabilization

Therefore, it can be shown that the voltages across R, namely, V_{R_a} and V_{R_b} at points a and b increase with the increase of frequency, R and C being kept constants. When the variation of the supply voltages V_{s_c} and V_{s_d} at points c and d with frequency is comparatively small and can be neglected (Figure 9), the values at normal frequency $\omega_0/2\pi$ will be used. The relation between V_R and V_s for $\omega_0/2\pi$ and for other two frequencies $\omega_1/2\pi$ and $\omega_2/2\pi$ slightly higher and lower than $\omega_0/2\pi$ respectively are shown in Figure 9.

The relation between the capacitance C and the amplitude of the flux linkage ϕ_m for $dV_R/dV_s=0$ and the relation between C and (ϕ_{m_c}, ϕ_{m_d}) are shown, Figure 14, the frequency $\omega/2\pi$ being constant; both relations are independent of R. ϕ_{m_c} and ϕ_{m_d} are the values at points c and d respectively.

From conditions of similarity, it can be shown that if the frequency increases from $\omega_2/2\pi$ to $\omega_1/2\pi$, the capacitance should be reduced by the factor ω_2^2/ω_1^2 to have the same values of ϕ_m.

Choice of R and C for Good Stabilization

Our choice is such that for certain ratio changes in the supply frequency and supply voltage $\pm K_f$ and $\pm K_s$ respectively, the stabilization in the load voltage is the best possible provided that the load voltage with respect to the supply voltage is reasonable. An approximate relation to satisfy these conditions is that the quantity Q given by

$$\{(V_{R_{a_1}}-V_{R_{b_2}})V_{s_{\text{mean}}}/2V_{R_{\text{mean}}}{}^2K_s=Q\}$$

is minimum where

$V_{s_{\text{mean}}}$ = mean supply voltage $=(V_{s_{c_0}}+V_{s_{d_0}})/2$

$V_{R_{\text{mean}}}$ = mean load voltage $=(V_{R_{a_1}}+V_{R_{b_2}})/2$

K_s = ratio change in the supply voltage $=(V_{s_{d_0}}-V_{s_{c_0}})/(V_{s_{d_0}}+V_{s_{c_0}})$

$V_{s_{c_0}}$ and $V_{s_{d_0}}$ are the supply voltages at points c_0 and d_0 of Figure 9, while $V_{R_{a_1}}$ and $V_{R_{b_2}}$ are the load voltages at points a_1 and b_2 respectively. The ratio of the supply voltages $V_{s_{c_0}}$ and $V_{s_{d_0}}$ at points c_0 and d_0 of Figure 9 is given by

$$V_{s_{d_0}}/V_{s_{c_0}}=(1+K_s)/(1-K_s)=K=\omega_0\phi_{m_{d_0}}\times\sqrt{(1+R/R')^2+R^2\{(1/\omega_0L_{eq_{d_0}})-\omega_0C\}^2}/\omega_0\phi_{m_{c_0}}\sqrt{(1+R/R')^2+R^2\{(1/\omega_0L_{eq_{c_0}})-\omega_0C\}^2}$$

The last relation gives

$$R=1/(\sqrt{M}-1/R')$$

where

$$M=[\{(1/\omega_0L_{eq_{c_0}})-\omega_0C\}^2(K\phi_{m_{c_0}}/\phi_{m_{d_0}})^2-\times\{(1/\omega_0L_{eq_{d_0}})-\omega_0C\}^2]/\{1-(K\phi_{m_{c_0}}/\phi_{m_{d_0}})^2\}$$

The indices 0, 1, and 2 refer to values at frequencies $\omega_0/2\pi$, $\omega_1/2\pi$, and $\omega_2/2\pi$ respectively.

For a certain value of C, $\phi_{m_{c_0}}$, $\phi_{m_{d_0}}$, and hence $L_{eq_{c_0}}$ and $L_{eq_{d_0}}$ can be obtained.

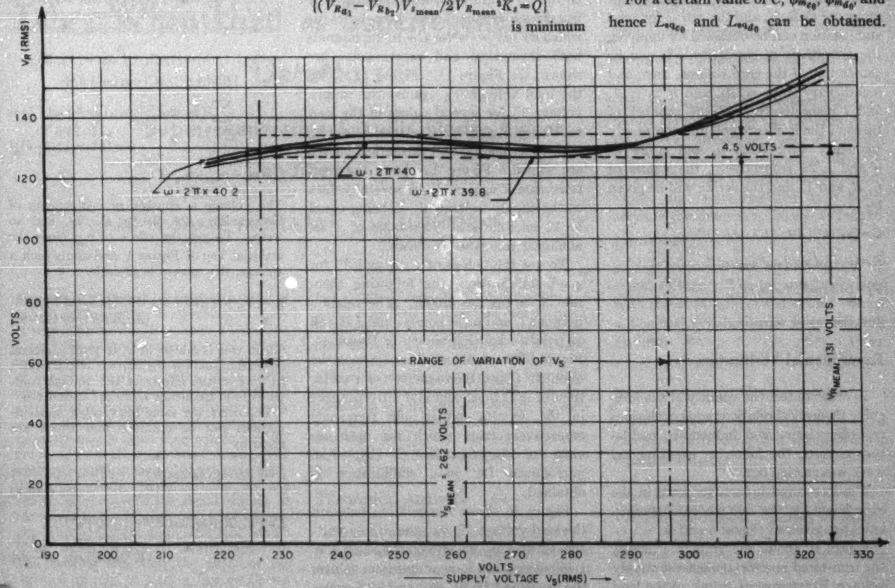


Figure 16. Voltage stabilizer characteristics

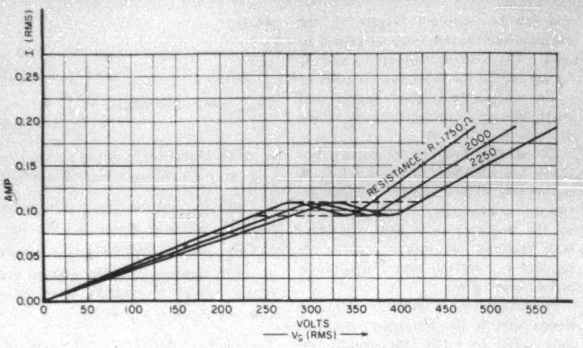


Figure 17. Current stabilizer characteristics

Substitution gives the corresponding value of R, from which the corresponding value of Q can be obtained. Similarly, for other values of the capacity C, the corresponding values of Q are obtained. The $Q \rightarrow C$ characteristic will show a minimum, giving the optimum capacity $C_{opt.}$ required for good stabilization, Figure 15, the corresponding values $R_{opt.}$ and $V_{s\,opt.}$ are consequently obtained.

For any other core of different dimensions, but of the same iron grade, the values of C, R, V_s and the rating for good stabilization can be obtained by applying the conditions of similarity. It can be shown that if the core volume, area, and length vary from τ_0, A_0, and l_0 to τ, A, and l respectively, and the number of turns of the iron-core coil and the supply frequency vary from n_0 and $\omega_0/2\pi$ to n and $\omega/2\pi$ respectively, the values of C, R, and V_s and the rating will be given by

$$C=(n_0^2/n^2)(A_0l/Al_0)(\omega_0^2/\omega^2)C_0$$

$$R=(l_0/l)(An^2/A_0n_0^2)(\omega/\omega_0)R_0$$

$$V_s=(nA/n_0A_0)(\omega/\omega_0)V_{s_0}$$

and the rating = (rating)$_0(\tau/\tau_0)(\omega/\omega_0)$.

Experimental Verification

In order to test the validity of the foregoing theoretical relations and conclusions regarding iron-cored inductance oscillatory circuits, the following experimental work was carried out.

The iron-cored inductance used in the experiments is the one whose characteristics are given in Figure 5 and 6.

Figure 10 shows the voltage V across the iron-cored reactor against the supply voltage V_s for several values of the load resistance R and for $C=10$ microfarads, both theoretical and experimental curves are indicated. As in section entitled "Some Properties of the Circuit," only the fundamental of the wave is used in the theoretical curves while rms values are used in the experimental curves. The value of V at nonlinear resonance, and which is independent of the load resistance R, is shown. The slopes dV_s/dV at such nonlinear points are also shown. The values of V and hence ϕ_m at which $dV/dV_s=$— are obtained experimentally from such curves. The boundaries at which $dV/dV_s=\infty$ for $C=10$ microfarads both theoretical and experimental are shown, see Figure 11. Figure 12 shows the load voltage V_R against the supply voltage V_s for several values of the load resistance R and for $C=10$ microfarads, both experimental and theoretical curves are shown. Figure 13 shows the load resistance R against V for several values of the supply voltage V_s, C being equal to 12 microfarads. The value of V at nonlinear resonance is shown.

To get the values of C, R and V_s for good stabilization, the following theoretical curves are drawn. The relation between C and ϕ_m at points c and d as well as points a and b is shown in Figure 14. Figure 15 shows the relation between the quantity Q and the capacity C for variations of ±12.5 per cent and ±0.5 per cent in the supply voltage and frequency respectively from which the optimum value of the capacity $C=C_{opt.}=7.25$ microfarads for good stabilization is obtained.

Figure 16 shows the relation between the load voltage V_R and the supply voltage V_s at $C=7.25$ microfarads and $R=$ 1,450 ohms, that is, near optimum values; the frequency $f=40$, 40.2 and 39.8 cycles per second. With these values, a stabilization as good as ±1.75 per cent in the load voltage or current which will be reduced at ±0.96 per cent if the frequency remains constant at 40 cycles per second; the supply voltage V_s varies between 227 and 297 volts.

Figure 17 shows the relation between the line current I and the supply voltage V_s for $C=8$ microfarads and for several values of the load resistance R and shows that good stabilization can be obtained with variable load resistance.

Appendix I. Determination of the Current *i* of Equation 5

According to the well-known "compensation theorem," if C_2 and C_{1_t} are short-circuited by closing the artificial key and replaced instead by a generator of zero internal impedance having an electromotive force $-\{(1/C_2)+(1/C_{1_t})\}\int i\,dt$, where i is the actual current flowing in the circuit, i would be unchanged; but the actual current flowing in the circuit is unknown; therefore, successive approximations will be adopted as follows

Closing the artificial key, the current would be given by

$$i_1=(V_s/R)\sin(\omega t+\theta)$$

i_1 gives the current to a first approximation, steady state only being considered.

For a second approximation, replace C_2 and C_{1_t} by a generator of zero internal impedance and having an electromotive force

$$\begin{aligned} v_1 = -\{(1/C_2)+(1/C_{1_t})\}\int i_1dt = (V_s/R)\times \\ [\{(2d+d^2)/\omega C\}\cos(\omega t+\theta)+ \\ \sum_{n=1}^{\infty}(h_{2_n}/2\omega C)\{\cos((2n+1)\omega t+\theta)+ \\ \cos((2n-1)\omega t-\theta)\}] \\ =v_{1_\omega}+v_{1_{3\omega}}+v_{1_{5\omega}}+\ldots \end{aligned}$$

The voltage v_1 consists of components of frequencies $\omega/2\pi$, $3\omega/2\pi$, $5\omega/2\pi$, and so forth. To get the effect of v_{1_ω}, close the artificial key of Figure 4, and apply such a voltage, the current is given by

$$i_2=(V_s/R)[\{(2d+d^2)/\omega C_1R\}\cos(\omega t+\theta)+ (h_2/2\omega CR)\cos(\omega t-\theta)]$$

When this is added to i_1, it gives a second approximation for the current in the circuit so long as the effects of the voltage components of frequency $\omega/2\pi$ are considered.

Following the same lines, other approximations can be obtained and which are given by

$$i_3=(V_s/R)[(h_2/2\omega CR)^2-\{(2d+d^2)/\omega C_1R\}^2] \sin(\omega t+\theta)$$

$$i_4=(V_s/R)[(h_2/2\omega CR)^2-[\{(2d+d^2)/ \omega C_1R\}^2]\{(2d+d^2)/\omega C_1R\}\cos(\omega t+\theta)+ (h_2/2\omega CR)\cos(\omega t-\theta)]$$

$$i_5=(V_s/R)[(h_2/2\omega CR)^2-\{(2d+d^2)/ \omega C_1R\}^2]^2\sin(\omega t+\theta)$$

Therefore, the fundamental component of the current, excluding the effects of the 3rd, 5th, . . . harmonic voltages, will be given by

$$\begin{aligned} i_{\text{Fund.}} &= i_1+i_2+i_3+i_4+i_5+\ldots \\ &= (V_s/R)[\sin(\omega t+\theta)+\{(h_2/2\omega CR)\times \cos(\omega t-\theta)+((2d+d^2)/\omega C_1R)\times \cos(\omega t+\theta)\}+\{(h_2/2\omega CR)^2- ((2d+d^2)/\omega C_1R)^2\}\sin(\omega t+\theta)+ \\ &\quad \{(h_2/2\omega CR)^2-((2d+d^2)/\omega C_1R)^2\}\times \{((2d+d^2)/\omega C_1R)\cos(\omega t+\theta)+ (h_2/2\omega CR)\cos(\omega t-\theta)\}+\{(h_2/2\omega CR)^2- ((2d+d^2)/\omega C_1R)^2\}^2\sin(\omega t+\theta)+\ldots] \\ &= (V_s/R)(1+a+a^2+a^3+\ldots)\times \{\sin(\omega t+\theta)+(h_2/2\omega CR)\cos(\omega t-\theta)+ ((2d+d^2)/\omega C_1R)\cos(\omega t+\theta)\} \\ &\simeq (V_s/R)\{\sqrt{M^2+N^2}/(1-a)\}\times \sin(\omega t+\theta_1) \end{aligned} \quad (11)$$

where

$$a=(h_2/2\omega CR)^2-\{(2d+d^2)/\omega C_1R\}^2$$

$$=(\omega L/R)^2[\{(h_2/2)(1+d)^2\}^2-(2d+d^2)^2]$$

$$M=\cos\theta+(\omega L/R)\{(h_2/2)(1+d)^2-(2d+d^2)\}\sin\theta$$

$$N=\sin\theta+(\omega L/R)\{(h_2/2)(1+d)^2+(2d+d^2)\}\cos\theta$$

and

$$\tan\theta_1=N/M$$

For a better approximation of $i_{\text{Fund.}}$, the effects of the components of the voltages v_3, v_5, et cetera of frequencies other than $\omega/2\pi$ would be considered. However, the approximation given by equation 11 is sufficient for the present nonlinear problem.

Harmonics other than the fundamental can be obtained using the method of successive approximation and following the same lines as before.

Therefore, the current i, to a satisfactory degree of approximation, will be as given in equation 6 in the text.

References

1. NON-LINEAR CIRCUITS FOR RELAY APPLICATION, **C. G. Suits.** *Electrical Engineering*, volume 50, December 1931, pages 963–65.

2. NON-LINEAR CIRCUITS APPLIED TO RELAYS, **C. G. Suits.** *Electrical Engineering*, volume 52, April 1933, pages 244–46.

3. CRITICAL CONDITIONS IN FERRORESONANCE, **P. H. Odessey, Ernst Weber.** *AIEE Transactions (Electrical Engineering)*, volume 57, August 1938, pages 444–52.

4. RESONANT NONLINEAR CONTROL CIRCUITS, **William T. Thomson.** *AIEE Transactions (Electrical Engineering)*, volume 57, August 1938, pages 469–76.

5. SIMILITUDE OF CRITICAL CONDITIONS IN FERRORESONANT CIRCUITS, **William T. Thomson.** *AIEE Transactions (Electrical Engineering)*, volume 58, March 1939, pages 127–30.

6. THE GENERALIZED SOLUTION FOR THE CRITICAL CONDITIONS OF THE FERRORESONANT PARALLEL CIRCUIT, **William T. Thomson.** *AIEE Transactions*, volume 58, 1939, pages 743–46.

7. AN UNSTABLE NONLINEAR CIRCUIT, **Claude M. Summers.** *AIEE Transactions (Electrical Engineering)*, volume 59, May 1940, pages 273–76.

8. MAGNETIC GENERATION OF A GROUP OF HARMONICS, **E. Peterson, J. M. Manley, L. R. Wrathall.** *Electrical Engineering*, volume 56, August 1937, pages 995–1001.

9. THE CLARIFICATION OF AVERAGE NEGATIVE RESISTANCE WITH EXTENSIONS OF ITS USE, **Brunetti.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 25, 1937, page 1595.

10. BESSEL FUNCTIONS FOR ENGINEERS (book), **N. W. McLachlan.** Oxford University Press, Oxford, England.

11. ELECTRON TUBE PERFORMANCE WITH LARGE APPLIED VOLTAGES, **Abd El-Samie Mostafa.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 39, 1951, page 70.

No Discussion

A New Method for Treating Electron Tubes When Used as Superregenerative Detectors

Part II. Superregenerative Circuits Under Signal Conditions

ABD EL-SAMIE MOSTAFA
NONMEMBER AIEE

M. EL-SHISHINI
NONMEMBER AIEE

Synopsis: This is the second of a series of papers dealing with a comprehensive study of superregenerative detection. In the first paper[1] the analysis under no-signal condition was given. In this paper, the analysis of superregenerative circuits under signal condition is presented. It is shown that the amplitude to which the oscillation under signal condition builds up, depends fundamentally on the quenching source and the nonlinearity of the tube characteristic. The automatic volume control action, the sensitivity, the selectivity, the stability, and the synchronization with the signal voltage or the semisynchronization (the synchronization with one of the equivalent side bands) of the signal voltage are clearly explained. The existing idea that the effect of the signal is merely to start a free oscillation with an amplitude depending on the signal is erroneous and does not conform with the fact. What actually happens when the signal is present, is that the system must, in general, be in some state of synchronization with the signal frequency or with one of the equivalent side bands. This latter is defined as semisynchronization. In order to have satisfactory reception, the free oscillation is generally suppressed. In general, no previous author has mentioned such synchronization or semisynchronization processes. Numerical examples are given for clarification.

Theoretical Analysis

Figure 1 shows a schematic diagram for a separately quenched superregenerative detector acted upon by an external source e of amplitude E and frequency $\omega_1/2\pi$ in series with the inductive branch.

Applying Kirchhoff's first law for the currents meeting at P we get

$$i_C+i_G+i_p+i_L=0$$

giving after differentiation

$$d^2v/dt^2+(1/C)(G\,dv/dt+di_p/dt)+\omega_0 v=\omega_0^2 e \quad (1)$$

where i_C, i_L, and i_G represent the alternating components of the currents in the capacitive, inductive, and conductive branches respectively, i_p, being the deviation of the plate current from its steady value, can be represented as a power series of the lumped voltage e' as follows

$$i_p=\alpha e'+\beta e'^2+\gamma e'^3$$

Near resonance, the amplitude E of the signal voltage will be too small compared with the amplitude of the output voltage v, therefore e' can be very approximately given by

$$e'=(1-\mu M/L)v+v_q$$

where v_q = quench voltage = $N\cos pt$

Differentiating i_p and neglecting dv_q/dt compared with dv/dt and substituting in equation 1 we get

---

Paper 53-157, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted March 10, 1952; made available for printing December 3, 1952.

ABD EL-SAMIE MOSTAFA is with the Alexandria University, Alexandria, Egypt. M. EL-SHISHINI is a member of the Egyptian Senate, Cairo, Egypt.

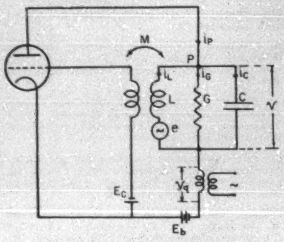


Figure 1. Separately quenched superregenerative detector with an applied signal

$$d^2v/dt^2+(1/C)\{G+\alpha(1-\mu M/L)+ 2\beta(1-\mu M/L)^2v+2\beta(1-\mu M/L)v_q+ 3\gamma(1-\mu M/L)^3v^2+6\gamma(1-\mu M/L)^2\times v_qv+3\gamma(1-\mu M/L)v_q^2\}dv/dt+\omega_0^2v =\omega_0^2e \quad (2)$$

As will be shown later in the experimental part (part III of this series[2]), the type of oscillation under signal condition depends on the amount of detune between $\omega_0/2\pi$ and $\omega_1/2\pi$ and on the amplitude E of the signal, where $\omega_0^2=1/LC$. If $\omega_0/2\pi$ is near $\omega_1/2\pi$, synchronization at the signal frequency takes place provided the conditions for stable synchronization are fulfilled; in this case, only synchronized oscillation of frequency $\omega_1/2\pi$ will be present, the free oscillation is suppressed. If $\omega_0/2\pi$ is near $(\omega_1\pm np)/2\pi$, synchronization at $(\omega_1\pm np)/2\pi$ takes place provided the conditions of the stability for such synchronization are fulfilled, n being an integer and $p/2\pi$ the quench frequency. Such synchronization does not take place at the signal frequency $\omega_1/2\pi$, but it takes place at an equivalent sideband frequency $(\omega_1\pm np)/2\pi$ and therefore it may be called semisynchronization. In this case, besides the semisynchronized oscillation of frequency $(\omega_1\pm np)/2\pi$, the tube also operates at frequency $\omega_1/2\pi$; however, free oscillation will be suppressed. In any case, such synchronization or semisynchronization is essential for satisfactory reception; the characteristic noise is generally suppressed. If the conditions of stable synchronization are not fulfilled, forced as well as free oscillations will be present, and in this case, the characteristic noise exists and the reception is poor. The two possible cases for satisfactory reception will now be treated.

SYNCHRONIZED OSCILLATION IF FREQUENCY $\omega_1/2\pi$ ONLY IS PRESENT, $\omega_0/2\pi$ IS NEAR $\omega_1/2\pi$

Assume v to be of the form $(V_t\cos\omega_1t+\ldots)$, where V_t is slowly varying compared with the frequency of oscillation. Therefore $dv/dt\approx -V_t\omega_1\sin\omega_1t+\ldots$ Let the signal voltage $e=E\cos(\omega_1t+\phi_t)$, then

$$\overset{*}{e}\approx k_tv\cos\phi_t+(1/\omega_1)(dv/dt)\sin\phi_t$$

Where $k_t=E/V_t$, it is a slow function of time, and where ϕ_t=the instantaneous phase angle between e and v, it is also a slow function of time. Substitution in equation 2 gives

$$d^2v/dt^2+(G_t/C)(1+h_{pt}\cos pt+h_{2pt}\times \cos 2pt+\ldots+h_{1t}\cos\omega_1t+h_{2t}\times \cos 2\omega_{1t}+\ldots)dv/dt+\omega_0^2\times (1-k_t\cos\phi_t)v=0 \quad (3)$$

where

$$G_t=G+\alpha(1-\mu M/L)-(\omega_0^2/\omega_1)Ck_t\sin\phi_t+ (3\gamma/2)(1-\mu M/L)^3V_t^2+(3\gamma/2)\times (1-\mu M/L)N^2$$

$$h_{pt}=(1/G_t)2\beta(1-\mu M/L)N$$

$$h_{2pt}=(1/G_t)(3\gamma/2)(1-\mu M/L)N^2$$

$$h_{1t}=(1/G_t)(2\beta+6\gamma N\cos pt)(1-\mu M/L)^2V_t$$

and

$$h_{2t}=(1/G_t)(3\gamma/2)(1-\mu M/L)^3V_t^2$$

In this way, the nonlinear differential equation 2 has been transformed to a linear one with periodically varying coefficients.

In the present analysis, as in the preceding paper,[1] it is assumed that the amount of nonlinearity is small, that is each of (G_t/ω_1C), $(G_t/\omega_1C)h_{pt}$, $(G_t/\omega_1C)h_{2pt}$ and so forth, and $(G_t/\omega_1C)h_{1t}$, $(G_t/\omega_1C)h_{2t}$, and so forth is much less than unity (this is equivalent to saying that the present work is limited to small signals).

Equation 3 is similar to the differential equation 3 of paper 1,[1] of the periodically varying resistance, with the substitution

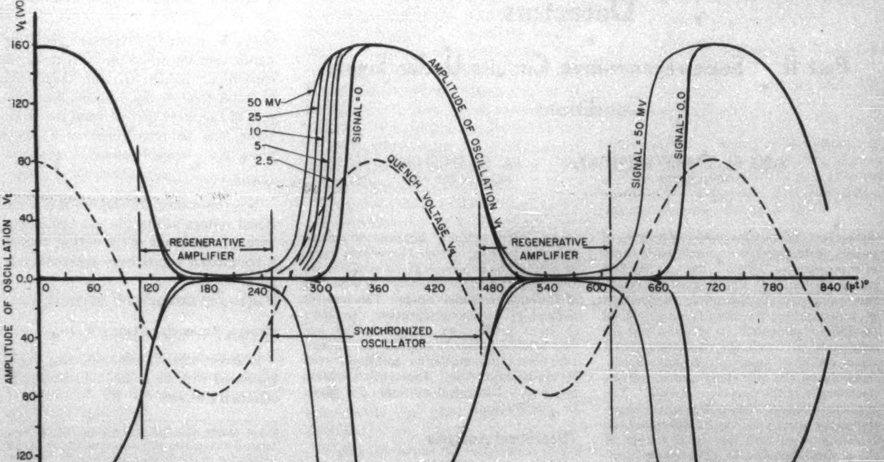


Figure 2. Envelopes of oscillations at no-signal and at signals equal to 2.5, 5.0, 10, 25, and 50 millivolts. Quench frequency = 100 kilocycles per second

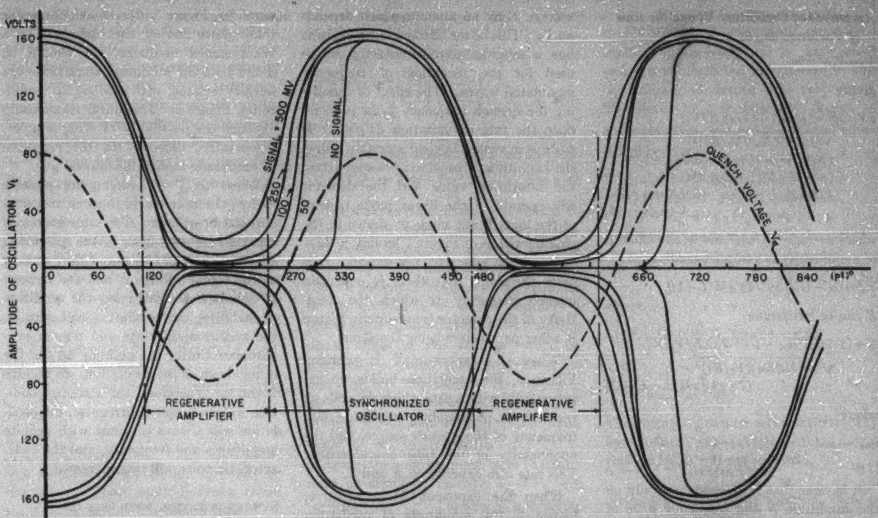


Figure 3. Envelopes of oscillations at no-signal and at signals equal to 50, 100, 250, and 500 millivolts. Quench frequency = 100 kilocycles per second

of v, G_t, L, and C for Q, R_{0l}, C, and L respectively. Also the assumptions made for the solution of the equation of the periodically varying resistance are satisfied in equation 3. However, the voltage v must be of the same form as assumed and which is $v = V_t \cos \omega_1 t + \ldots$ Therefore a possible solution for equation 3 can be obtained from equation 7 of paper 1[1] provided $\phi' - \theta' = \pi/2$ that is $d \approx 0$; therefore

$$v = V_s e^{\int X dt} \{\cos \omega_1 t - (1/6\omega_1 C)(2\beta - 6\gamma' N \times \cos pt)(\mu M/L - 1)^2 V_t \sin 2\omega_1 t - (1/8\omega_1 C)(3\gamma'/4)(\mu M/L - 1)^3 V_t^2 \times \sin 3\omega_1 t - \ldots\} \quad (4)$$

where

$$X = (1/2C)\{\alpha' - G + \omega_1 C k_t \sin \phi_t - (3\gamma'/4) \times (\mu M/L - 1)^3 V_t^2 - (3\gamma'/2) \times (\mu M/L - 1) N^2 + (\mu M/L - 1) \times (2\beta N \cos pt - (3\gamma'/2) N^2 \cos 2pt)\} \quad (4A)$$

$$\omega_1^2 \approx \omega_0^2 \{1 - (E/V_t) \cos \phi_t\} \quad (5)$$

$$\gamma' = -\gamma, \; \alpha' = \alpha(\mu M/L - 1)$$

and

$$V_t = V_s e^{\int X dt}$$

giving the steady-state solution, there is no transient. V_s is the amplitude of oscillation at the instant at which X just starts to be positive. From equation 5 $\cos \phi_t = \{1 - \omega_1^2/\omega_0^2\} V_t/E$, showing that for ω_1 slightly less than ω_0, ϕ_t varies from slightly less than $\pi/2$ to slightly higher than zero as V_t increases; for ω_1 slightly higher than ω_0, ϕ_t varies from slightly higher than $\pi/2$ to slightly less than π as V_t increases. Now

$$\sin \phi_t \approx \sqrt{1 - (Z/\omega_1)^2/(E/V_t)^2}$$

where Z = amount of detune $= 2(\omega_0 - \omega_1)$; substitution in equation 4(A) gives

$$X \approx (1/2C)[\alpha' - G + \omega_1 C(E/V_t) \times \sqrt{1 - (Z/\omega_1)^2/(E/V_t)^2} - (3\gamma'/4)(\mu M/L - 1)^3 V_t^2 - (3\gamma'/2)(\mu M/L - 1) N^2 + (\mu M/L - 1) \times \{2\beta N \cos pt - (3\gamma'/2) N^2 \cos 2pt\}]$$

The frequency of oscillation is that of the signal voltage, but the phase angle between the signal voltage e and the output voltage v and which is ϕ_t is varying according to equation 5; the phase angle ϕ_t is modulated with a fundamental frequency = frequency of quench source = $p/2\pi$.

Again, as in the first paper,[1] when there is no signal applied, multiple resonance phenomena exists, but in this case, the equivalent side bands are of frequencies $(\omega_1 + np)/2\pi$ and $n = \pm$integer.

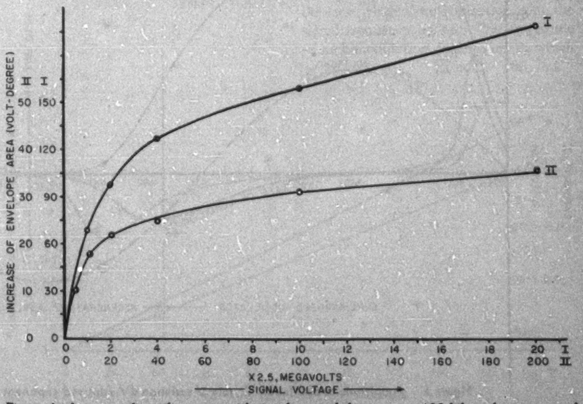


Figure 4. Increase of envelope area with signal, quench frequency = 100 kilocycles per second

Amplitude of Oscillation V_t and the Rate of Variation dv_t/dt

$$V_t = V_s \epsilon^{\int X dt}$$

giving

$$dV_t/dpt = (\epsilon/2)[1 + (1/\epsilon)(E/V_t)\sqrt{1-\{(Z/\omega_1)(V_t/E)\}^2} - (V_t^2/V_0^2) - \{(3\gamma'/2)(\mu M/L-1)/(\alpha'-G)\}N^2(1+\cos 2pt) + \{2\beta(\mu M/L-1)/(\alpha'-G)\}N\cos pt] \times V_t(\omega_1/p) \quad (6)$$

where $\epsilon = (\alpha' - G)/\omega_1 C$ and as the amount of nonlinearity is assumed small, $\epsilon << 1$

$$V_s^2 = (\alpha' - G)/(3\gamma'/4)(\mu M/L-1)^3$$

X can be written as

$$X = (1/2C)\{(\alpha_{eq}' - G) + \omega_1 C(E/V_t) \times \sqrt{1-\{(Z/\omega_1)(V_t/E)\}^2} - (3\gamma'/4)(\mu M/L-1)^3 V_t^2\}$$

where

$$\alpha_{eq}' = (\mu M/L-1)\{(1/\rho) - (3\gamma'/2)N^2 + 2\beta N\cos pt - (3\gamma'/2)N^2\cos 2pt\}$$

It is clear that X depends not only on the amplitude N and frequency $p/2\pi$ of the quench source and the nonlinearity of the tube characteristic and the amplitude of oscillation V_t, but also on the signal voltage E and the amount of detune Z. The effect of the signal is equivalent to a reduction in the load conductance G, the amount of reduction depends on V_t. The effect of the detune Z is equivalent to a reduction in the signal voltage E by an amount which depends on V_t. This latter characteristic explains how a superregenerative detector can be used for the reception of frequency modulated waves. The effect of increasing the quench frequency $p/2\pi$ is to reduce the rate of variation dV_t/dpt. If $p/2\pi$ is increased beyond a certain value, the amplitude of oscillation cannot attain the saturation value and the detector will operate in the linear mode instead of the logarithmic mode as shown in the characteristics of Figure 7 having $p/2\pi =$ 200 and 300 kilocycles per second. It is clear therefore, that there is a certain quench frequency at which the sensitivity of the detector is maximum, Figure 8, other parameters being the same.

When the exponent X is positive. Figure 5, the oscillation builds up exponentially with a rate which depends on the amount of coupling, amplitude, and frequency of the quench source, and the nonlinearity of the tube characteristic. This rate also depends on E and Z.

When the exponent X is negative, Figure 5, the amplitude of oscillation decreases. During the interval at which $(\alpha_{eq}' - G)$ is positive, the system works as a synchronized oscillator with variable regeneration. When $(\alpha_{eq}' - G)$ is negative, the system works as a regenerative amplifier whose regeneration varies as the amplitude of $(\alpha_{eq}' - G)$ varies, Figures 2, 3, and 5. From the relation $V_t = V_s \epsilon^{\int X dt}$, it can easily be shown that V_t is maximum or minimum when $X = 0$. The minimum value occurs during the operation as an amplifier, Figure 5. The maximum value takes place during the operation as a synchronized oscillator. It should be noted that the minimum amplitudes are linearly related with the signal amplitudes, Figure 6. This linear relationship explains the possibility of using a superregenerative detector for the reception of amplitude modulated waves.

When $(\alpha_{eq}' - G)$ enters the positive region, the tube starts to work as a synchronized oscillator. The same process is repeated for each cycle of the quenching source. It should be noted that so long as the signal voltage E and the amount of detune Z are satisfying the condition of stability, the oscillation will start in the synchronized state and remain synchronized during the building up and the dying down of the oscillation, see Figures 2 and 3. Therefore, the starting condition will no longer be irregular, the oscillation will always continue with definite amplitudes and frequency, and the characteristic noise will be suppressed.

SYNCHRONIZATION WITH ONE OF EQUIVALENT SIDE BANDS OR SEMISYNCHRONIZATION

As mentioned before, if the free oscillation frequency $\omega_0/2\pi$ is near $(\omega_1 + np)/2\pi$, where $n = \pm$integer, synchronization at $(\omega_1 + np)/2\pi$ takes place provided the conditions of stability for such synchronization are fulfilled. As this synchronization does not take place at the signal frequency $\omega_1/2\pi$ but it takes place at an equivalent side-band frequency

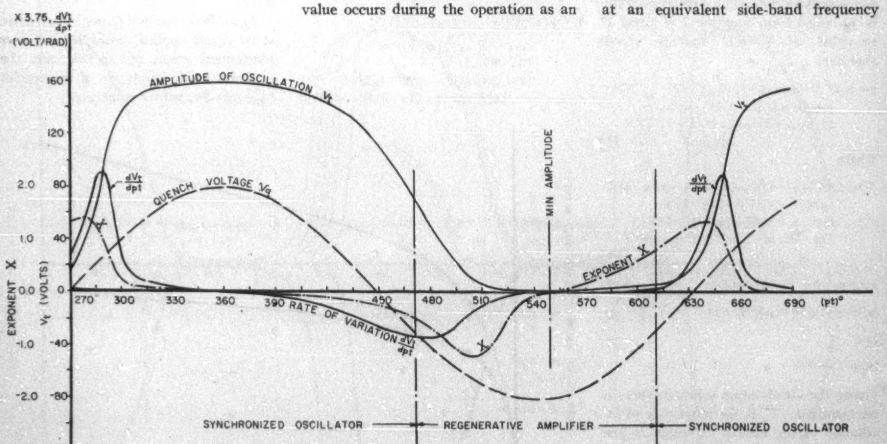


Figure 5 Amplitude of oscillation V_t, rate of variation dV_t/dpt and exponent X, quench frequency = 100 kilocycles per second. Signal voltage = 50 millivolts

$(\omega_1+np)/2\pi$, therefore, it may be termed semisynchronization. The output voltage will consist of two components, namely, the forced oscillation v_1 of frequency $\omega_1/2\pi$ and the semisynchronized oscillation v_{0_1} which can also be considered forced, where $v_1=V_{t_1}\cos\omega_1 t+\ldots$ and $v_{0_1}=V_{t_{0_1}}\cos(\omega_1+np)t+\ldots$, and zero time is chosen such that the phase angle between v_1 and v_{0_1} is zero; $v=v_1+v_{0_1}$. Therefore, in this case also the free oscillation will be suppressed; the semisynchronized oscillation of frequency $(\omega_1+np)/2\pi$ which is near $\omega_0/2\pi$ will exist instead. The amplitudes V_{t_1} and $V_{t_{0_1}}$ are slow functions of time. The signal voltage

$$e=E\cos(\omega_1 t+\phi_t)=E\cos\{(\omega_1+np)t+(\phi_t-npt)\}=k_t\{V_{t_{0_1}}\cos(\phi_t-npt)+(dv_{0_1}/dt)(1/(\omega_1+np))\sin(\phi_t-npt)\}$$

where

$$k_t=E/V_{t_{0_1}}$$

is a slow function of time; $\phi_t-npt=$ instantaneous phase angle between e and v_{0_1}, it is also a slow function of time.

In order to simplify the analysis and not to lose its generality, the plate current i_p will be taken as $i_p=\alpha' e+\gamma e'^3$, in which case all terms containing β will be omitted. Again, substitution of $v=v_1+v_{0_1}=V_{t_1}\cos\omega_1 t+V_{t_{0_1}}\cos(\omega_1+np)t$ in the coefficient of dv/dt in equation 2 gives

$$\begin{aligned}&[d^2v_{0_1}/dt^2+(1/C)[G-\alpha'-(\omega_0^2/(\omega_1+np))\times\\&\quad Ck_t\sin(\phi_t-npt)+3\gamma'(\mu M/L-1)^3\times\\&\quad\{V_{t_1}\cos\omega_1 t+V_{t_{0_1}}\cos(\omega_1+np)t\}^2+\\&\quad 3\gamma'(\mu M/L-1)N\cos pt\{V_{t_1}\cos\omega_1 t+\\&\quad V_{t_{0_1}}\cos(\omega_1+np)t\}+(3\gamma'/2)\times\\&\quad(\mu M/L-1)N^2(1+\cos 2pt)]dv_{0_1}/dt+\\&\quad\omega_0^2\{1-k_t\cos(\phi_t-npt)\}v_{0_1}]+\\&\quad[d^2v_1/dt^2+(1/C)[G-\alpha'+3\gamma'\times\\&\quad(\mu M/L-1)^3\{V_{t_1}\cos\omega_1 t+V_{t_{0_1}}\cos(\omega_1+\\&\quad np)t\}^2+3\gamma'(\mu M/L-1)N\cos pt\times\\&\quad\{V_{t_1}\cos\omega_1 t+V_{t_{0_1}}\cos(\omega_1+np)t\}+\\&\quad(3\gamma'/2)(\mu M/L-1)N^2(1+\cos 2pt)]\times\\&\qquad dv_1/dt+\omega_0^2 v_1]=0\end{aligned}\qquad(7)$$

Equation 7 is composed of two differential equations, the first one when equated to zero gives the semisynchronized oscillation v_{0_1} while the second one when equated to zero gives the oscillation v_1.

Therefore, the semisynchronized oscillation is represented by the equation

$$\begin{aligned}&d^2v_{0_1}/dt^2+(G_t/C)[1+h_{2p_t}\cos 2pt+h_{np_t}\times\\&\quad\cos npt+h_{1_t}\cos\{(\omega_1+np)t-\theta_{1_t}\}+h_{2_t}\\&\quad\cos\{2(\omega_1+np)t-\theta_{2_t}\}]dv_{0_1}/dt+\omega_0^2\times\\&\qquad\{1-k_t\cos(\phi_t-npt)\}v_{0_1}=0\end{aligned}\qquad(8)$$

where

$$\begin{aligned}G_t=&\{G-\alpha'-(\omega_0^2/(\omega_1+np))Ck_t\\&\sin(\phi_t-npt)+(3\gamma'/2)(\mu M/L-1)^3\times\\&(V_{t_1}^2+V_{t_{0_1}}^2)+(3\gamma'/2)(\mu M/L-1)N^2\}\end{aligned}$$

$$h_{2p_t}=(1/G_t)(3\gamma'/2)(\mu M/L-1)N^2$$

$$h_{np_t}=(1/G_t)(3\gamma')(\mu M/L-1)^3V_{t_1}V_{t_{0_1}}$$

$$h_{2_t}=(1/G_t)(3\gamma'/2)(\mu M/L-1)^3\times(V_{t_1}^2+V_{t_{0_1}}^2+2V_{t_1}V_{t_{0_1}}\cos npt)$$

and

$$\tan\theta_{2_t}=\{V_{t_1}^2\sin 2npt+2V_{t_1}V_{t_{0_1}}\sin npt\}/\times\{V_{t_1}^2\cos 2npt+V_{t_{0_1}}^2+2V_{t_1}V_{t_{0_1}}\cos npt\}$$

The differential equation 8 is similar to equation 3 of paper 1,[1] of a series oscillatory circuit having a periodically variable resistance. The only difference is that the component of frequency $2(\omega_1+np)/2\pi$ in the coefficient of dv_{0_1}/dt contains a phase angle θ_{2_t} and therefore the resulting phase angle in the solution must be modified; a phase angle $\theta_{2_t}/2$ must be added to it. The solution can therefore be obtained from equation 7 of paper 1, putting $h_{1_t}=0$ and modifying the phase angle $(\phi'-\theta')$ to $\{(\theta_{2_t}/2)+\phi'-\theta'\}$.

Again, the solution must be of the same form as assumed, namely, $v_{t_{0_1}}=V_{t_{0_1}}\cos(\omega_1+np)t+\ldots$. Therefore, a possible solution for equation 8 can be obtained from that of periodically variable resistance oscillatory circuits provided

$$\{(\theta_{2_t}/2)+(\phi'-\theta')\}=\pi/2$$

giving

$$\begin{aligned}&(1/2)\tan^{-1}\{(V_{t_1}^2\sin 2npt+2V_{t_1}V_{t_{0_1}}\times\\&\quad\sin npt)/(V_{t_1}^2\cos 2npt+V_{t_{0_1}}^2+2V_{t_1}\times\\&\quad V_{t_{0_1}}\cos npt)\}=-\tan^{-1}[2(2d+d^2)/\times\\&\quad\{(4\sqrt{a_t})/(\omega_1+np)+h_{2_t}G_t/\times\\&\quad(\omega_1+np)C\}]+\tan^{-1}\{(-\sqrt{a_t})/\times\\&\qquad(\omega_1+np)\}\end{aligned}\qquad(9)$$

where

$$\sqrt{a_t}=\{(\omega_1+np)/4\}[\{h_{2_t}G_t/(\omega_1+np)C\}^2-\{2(2d+d^2)\}^2]^{1/2}$$

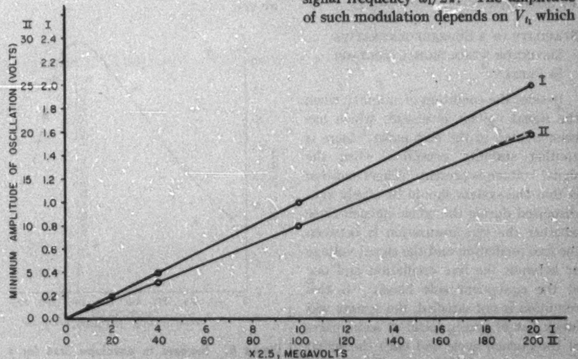


Figure 6. Minimum amplitude of oscillation versus signal voltage

As the amount of nonlinearity is assumed small, $\sqrt{a_t}/(\omega_1+np)$ will be very small and can be neglected in equation 9; also the amount of detune d is very small. Therefore, to a satisfactory degree of approximation, $\sqrt{a_t}\approx h_{2_t}G_t/4C$ and equation 9 becomes

$$\begin{aligned}&-(2d+d^2)=1-\{\omega_{0_1}^2/(\omega_1+np)^2\}\\&\quad=\{3\gamma'(\mu M/L-1)^3/4(\omega_1+np)C\}\times\\&\quad\{V_{t_1}^2+V_{t_{0_1}}^2+2V_{t_1}V_{t_{0_1}}\cos npt\}\times\\&\quad(1/2)\{(V_{t_1}^2\sin 2npt+2V_{t_1}V_{t_{0_1}}\sin npt)/\times\\&\quad(V_{t_1}^2\cos 2npt+V_{t_{0_1}}^2+2V_{t_1}V_{t_{0_1}}\times\\&\qquad\cos npt)\}\end{aligned}\qquad(10)$$

where

$$\omega_{0_1}^2\approx\omega_0^2\{1-(E/V_{t_{0_1}})\cos(\phi_t-npt)\}$$

and

$$(1+d)=\omega_{0_1}/(\omega_1+np)$$

The voltage v_{0_1} can be given satisfactorily by

$$v_{0_1}=V_s e^{\int X_1 dt}\{\cos(\omega_1+np)t+\ldots\}\qquad(11)$$

giving the steady-state solution, there is no transient, where

$$\begin{aligned}X_1\approx&(1/2C)\{\alpha'-G+(\omega_0^2C/(\omega_1+np))\times\\&(E/V_{t_{0_1}})\sin(\phi_t-npt)-(3\gamma'/4)\times\\&(\mu M/L-1)^3(V_{t_1}^2+V_{t_{0_1}}^2)-\\&(3\gamma'/2)(\mu M/L-1)N^2-(3\gamma'/2)\times\\&(\mu M/L-1)N^2\cos 2pt-(3\gamma'/2)\times\\&\quad(\mu M/L-1)^3V_{t_1}V_{t_{0_1}}\cos npt\}\end{aligned}\qquad(12)$$

V_s in equation 11 is the amplitude of oscillation at the instant at which X_1 just starts to be positive.

From relations 10 and 12, $\cos(\phi_t-npt)$ and hence $\sin(\phi_t-npt)$ can be obtained.

It is shown from the solution of v_{0_1} that its envelope is modified; it is modulated with fundamental frequency $np/2\pi$. This is besides the modulation appearing in the case where synchronization is at the signal frequency $\omega_1/2\pi$. The amplitude of such modulation depends on V_{t_1} which

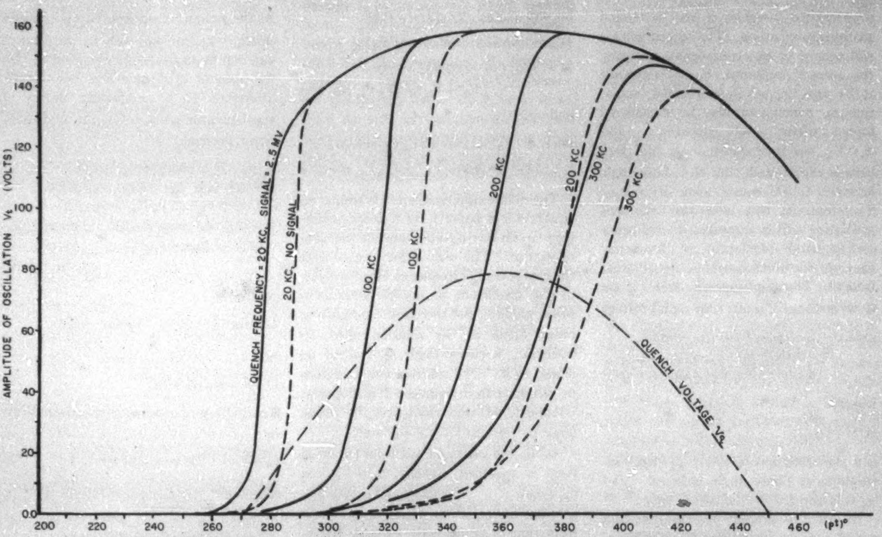


Figure 7. Building up curves for different quench frequencies at no-signal and at a signal of 2.5 millivolts

latter depends on the amount of detune between ω_1 and ω_0. As n increases, V_{t_1} is reduced and hence the amplitude of such modulation is reduced.

The oscillation v_1 is represented by the second differential equation of 7. As far as v_1 is concerned, the tube works as a detuned amplifier. This detuned amplification is also affected by the "semi-synchronized oscillation" v_{0_1} as seen from the differential equation. The amplitude of v_1 depends on the amount of detune between $\omega_1/2\pi$ and $\omega_0/2\pi$. As n increases, the amount of detune increases and hence V_{t_1} is reduced.

STABILITY OF A SUPERREGENERATIVE DETECTOR WHEN SIGNAL VOLTAGE IS PRESENT

Besides the condition of stability when the signal voltage is absent, which has been treated in the first paper,[1] there is another stability condition when the signal voltage is present. This condition is that the system should be stably synchronized during the whole quench cycle whether the synchronization is between the free oscillation and the signal voltage or between the free oscillation and one of the equivalent side bands. If this condition is not satisfied, the system will break out of synchronism at some parts of the quench cycle and both the signal frequency and the free oscillation frequency will be present, the initial conditions at the start of the oscillations will be irregular and the characteristic noise will exist.

Following the same lines as in the case when the signal voltage is absent, and which was treated in the first paper,[1] the condition of stability of a synchronized superregenerative detector can be obtained. The two cases will be treated separately.

Synchronization at Signal Frequency

The condition for stability is $(dX/dV_m)_{V_t=V_m}<0$. Using equation 4(A) we get

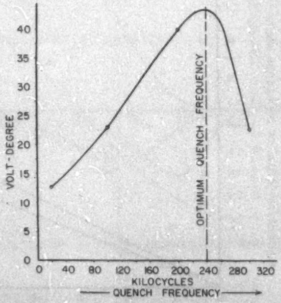


Figure 8. Increase in envelope area for a constant signal (=2.5 millivolts) versus quench frequency

$$(d/dV_m)\{\alpha'-G+\omega_1Ck_{t_m}\sin\phi_{t_m}-(3\gamma'/4)(\mu M/L-1)^3V_m^2-(3\gamma'/2)\times(\mu M/L-1)N^2\}<-(d/dV_m)\times\{2\beta(\mu M/L-1)N\cos pt_m-(3\gamma'/2)\times(\mu M/L-1)N^2\cos 2pt_m\}$$

where ϕ_{t_m} is the phase angle at $V_t=V_m=$ maximum amplitude, $k_{t_m}=E/V_m$, and t_m is the time at $V_t=V_m$.

If the quench source were absent, the system would act as a simple synchronized oscillator and if its stationary amplitude $=V_m$, then its condition of stability would be

$$(d/dV_m)\{\alpha'-G+\omega_1C(E/V_m)\sin\phi_{t_m}-(3\gamma'/4)(\mu M/L-1)^3V_m^2\}<0$$

Therefore, a stable synchronized system in the superregenerative state may be unstable in the simple synchronized state for the same maximum amplitude of oscillation.

Synchronization at an Equivalent Side-Band Frequency

The condition for stability is $(dX_1/dV_{0_{1m}})<0$; using equation 12 we get

$$(d/dV_{0_{1m}})\{\alpha'-G+(\omega_0^2C/(\omega_1+np))k_{t_m}\times\sin(\phi_{t_m}-npt_m)-(3\gamma'/4)(\mu M/L-1)^3(V_{t_{01m}}^2+V_{t_{1m}}^2)-(3\gamma'/2)\times(\mu M/L-1)N^2\}<(d/dV_{0_{1m}})\times\{(3\gamma'/2)(\mu M/L-1)^3V_{t_{01m}}V_{t_{1m}}\times\cos npt_m+(3\gamma'/2)(\mu M/L-1)N^2\times\cos 2pt_m\}$$

where $V_{t_{1m}}$ is the amplitude of V_{t_1} when

$V_{t_{01}}$ is maximum (that is, $dV_{t_{01}}/dt=0$) of amplitude $V_{t_{01m}}$, $k_{tm}=E/V_{t_{01m}}$, and t_m is the time at which $V_{t_{01}}$ is maximum.

This shows that the condition of stable synchronization at an equivalent side-band frequency depends on n, that is on the order of the side band, as well as on the amplitude $V_{t_{1m}}$. The latter depends also on the amount of detune between $\omega_0/2\pi$ and $\omega_1/2\pi$, that is, on the order of the side band. Therefore, there is a maximum value for n at which stable synchronization ceases.

Numerical Examples

From the foregoing theoretical analysis, most of the phenomenon of superregenerative receiving circuits can be obtained. The sensitivity, the automatic volume control action, the optimum quench frequency, and the different operating states during a quench cycle can be clarified by the following numerical examples. To simplify matters, the examples will be taken for the case when the signal frequency $=\omega_0/2\pi$, stably synchronized oscillation being assumed.

Let the signal frequency $=\omega_1/2\pi=75$ megacycles per second

$\epsilon=(\alpha'-G)/\omega C=0.01$

$V_0=100$ volts

$\{(3\gamma'/2)(\mu M/L-1)/(\alpha'-G)\}N^2=0.5$

$Z=0$

and

$\{2\beta(\mu M/L-1)/(\alpha'-G)\}N=2.5$

Substitution in equation 6 gives

$$dV_t/dpt=\{3.75\times10^6/(p/2\pi)\}\times \{0.5(1+5\cos pt-\cos 2pt)+100(E/V_t)-(V_t^2/100^2)\}V_t \quad (13)$$

The amplitude V_t can be obtained graphically using equation 13 for any suitable values of E and $p/2\pi$. The graphical integration is carried out by the step-by-step method. For a quench frequency of 100 kilocycles per second, the envelopes of oscillation at no-signal and at signals equal to 2.5, 5.0, 10.0, 25.0, and 50.0 millivolts are shown in Figure 2. The regions in which the system operates as a regenerative amplifier and a synchronized oscillator are clearly shown. It is clear from Figure 2 that the maximum amplitude of oscillation can be considered independent of the signal. However, if the signal voltage is higher than 100 millivolts, the maximum amplitude of the output voltage will depend upon the signal, Figure 3. Figure 4 shows the increase in envelope area with signal indicating the limiting action of superregenerative detectors, quench frequency being 100 kilocycles per second. Figure 5 gives the rate of variation dV_t/dpt, the exponent X, the amplitude of oscillation V_t, and the quench voltage v_q, signal amplitude = 50 millivolts, and quench frequency = 100 kilocycles per second; the position of the minimum value of V_t is clearly shown. It should be noted that V_t is minimum at $X=0$. At the minimum value of V_t along a quench cycle, the system works as a linear amplifier; Figure 6 shows clearly the linear relationship between the minimum value of V_t and the signal amplitude, quench frequency being 100 kilocycles per second.

To show that there is an optimum quench frequency at which the sensitivity is maximum, several envelopes are calculated under no-signal and under signal conditions for quench frequencies of 20, 100, 200, and 300 kilocycles per second, see figure 7. It is clear that at the quench frequencies 200 and 300 kilocycles per second, the detector operates in the linear mode. At 20 and 100 kilocycles per second, it operates in the logarithmic mode. Figure 8 shows the increase in envelope area for a constant signal (being proportional to the sensitivity) against quench frequency, the existence of an optimum quench frequency is clearly indicated.

References

1. A New Method for Treating Electron Tubes when Used as Superregenerative Detectors. Part I—Superregenerative Circuits Under No-Signal, **Abd El-Samie Mostafa, M. El-Shishini.** *AIEE Transactions*, volume 72, part I, May 1953, pages 207–13.

2. A New Method for Treating Electron Tubes when Used as Superregenerative Detectors. Part III—Experimental Investigation of Superregenerative Circuits, **Abd El-Samie Mostafa, M. El-Shishini.** *AIEE Transactions*, volume 72, part I, July 1953, pages 290–97.

3. Electronics Manual for Radio Engineers (book). McGraw-Hill Book Company, Inc., New York, N. Y., "Superregenerative Detection Theory," W. F. Bradley, 1949, page 606.

4. *Ibid.*, "Superregenerator Design," A. Hazeltine, D. Richman, B. D. Loughlin, page 609.

5. *Ibid.*, "Superregenerative Detector Selectivity," A. Easton, page 613.

6. Radio Engineering (book), F. E. Terman. McGraw-Hill Book Company, Inc., New York, N. Y., 1947.

No Discussion

A New Method for Treating Electron Tubes When Used as Superregenerative Detectors

Part III. Experimental Investigation of Superregenerative Circuits

ABD EL-SAMIE MOSTAFA
NONMEMBER AIEE

M. EL-SHISHINI
NONMEMBER AIEE

Synopsis: This is the third of a series of papers dealing with a comprehensive study of superregenerative detection. In the first paper[1] the analysis under no-signal condition was given. In the second paper[2] the analysis under signal condition was presented. In the present paper, an experimental investigation on superregenerative receiving circuits is carried out to clarify most of the phenomenon involved in these circuits, among which are: the selectivity, the sensitivity, the optimum quench frequency, and the most important phenomena, namely, the synchronization with the signal voltage at the signal frequency $\omega_1/2\pi$ and the semisynchronization or the synchronization with one of the equivalent side bands (that is, the synchronization at a frequency equal to $(\omega_1 \pm np)/2\pi$, $p/2\pi=$ quench frequency). In this way, the operating mechanism of superregenerative detectors is clearly explained. For proper reception, free oscillations are generally suppressed. The experimental results are in good agreement with the theoretical predictions given in the preceding papers.[1,2]

Owing to apparatus limitations, it was considered more appropriate to limit the work to low-frequency operation, the quench frequency being suitably chosen so that a proper relation between operating and quench frequencies is always maintained. Two different operating frequencies were chosen, one being in the neighbourhood of 100 kilocycles per second, the other being around 1,000 kilocycles per second. The first frequency is used to demonstrate a certain set of characteristics of superregenerative detectors, while the second frequency is used to demonstrate a different set of characteristics.

Paper 53-155, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted March 10, 1952; made available for printing December 4, 1952.

Abd El-Samie Mostafa is with the Alexandria University, Alexandria, Egypt. M. El-Shishini is a member of the Egyptian Senate, Cairo, Egypt.

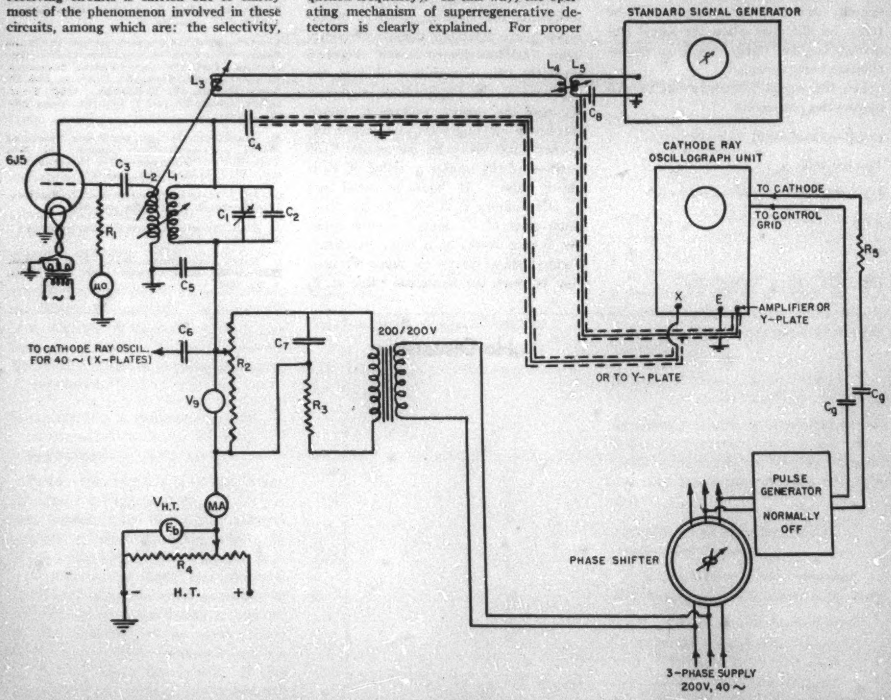
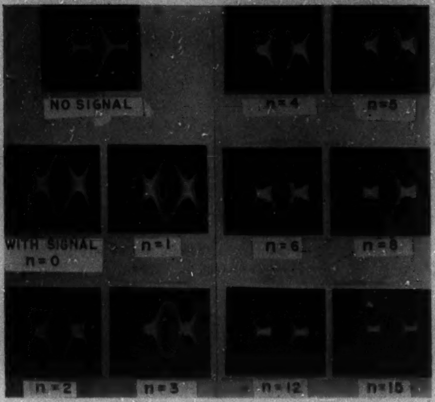


Figure 1. Connection diagram, operating frequency=100 kilocycles per second. $L_1=1{,}560$, $L_2=1{,}560$, $L_3=160$, $L_4=1{,}045$, $L_5=1{,}560$, $C_1=1{,}200$, $C_2=1{,}500$, $C_3=500$, $C_4=100$, $C_5=10{,}000$, $C_6=0.25$ microfarad, $C_7=40$ microfarads, $C_8=2{,}500$, $C_9=0.01$ microfarad (2,500-V), $R_1=50{,}000$, $R_2=690$, $R_3=355$, $R_4=5{,}500$, $R_5=200{,}000$; inductances are in microhenrys, capacitances in micromicrofarads (except when stated), and resistances in ohms


NO SIGNAL
n=4
n=5
WITH SIGNAL
n=0
n=1
n=6
n=8
n=2
n=3
n=12
n=15

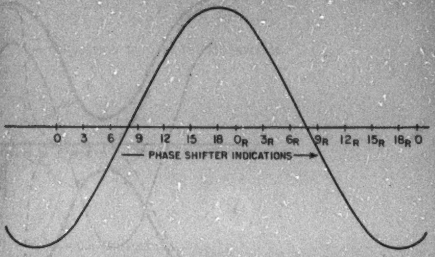


Figure 2 (left). Several oscillograms at a constant high signal and for several values of n, where $f_1 \cong f_0 - np/2\pi$, f_1=signal frequency, f_0=free oscillation frequency, and $p/2\pi$=quench frequency

Figure 4 (above). Calibrated quench cycle, giving indications of phase shifter against the instant at which the electron beam of the cathode-ray oscillograph appears on the screen. The index R indicates that the connection to the phase shifter is reversed

Experimental Characteristics of Superregenerative Detectors when Signal Frequency Is About 100 Kilocycles per Second

Figure 1 represents a complete connection diagram; the values of the different elements are indicated. The d-c plate voltage is taken 120 volts, while the quench voltage is 40 volts (rms). The coupling between plate and grid coils is so adjusted that free oscillation just starts in the absence of the quench source.

The output voltage is connected, through a 100-micromicrofarad coupling capacitor to the *Y*-plates of a cathode-ray oscillograph; the time base frequency is synchronized to the quench frequency. With no signal applied, the oscillations are not absolutely steady; the envelope during the building up of the oscillation

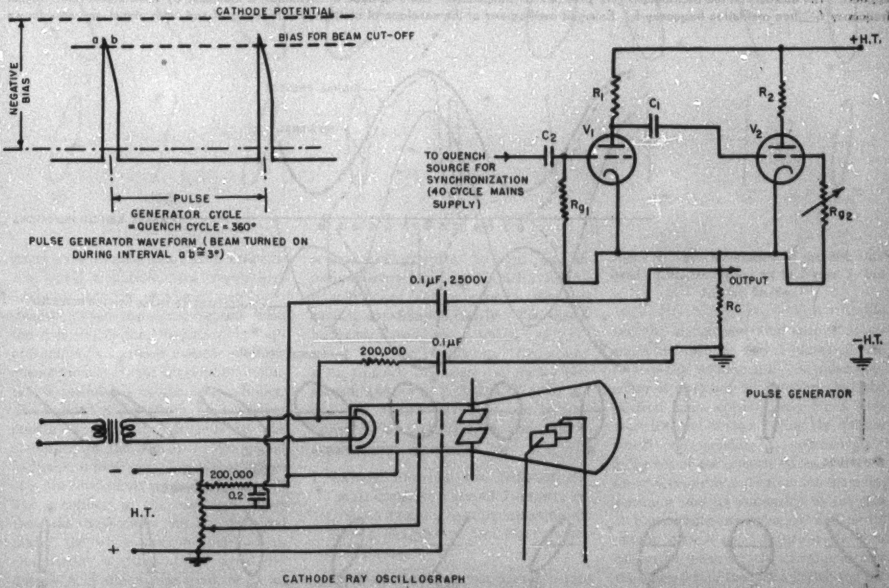

CATHODE POTENTIAL
BIAS FOR BEAM CUT-OFF
NEGATIVE BIAS
a b
PULSE
GENERATOR CYCLE = QUENCH CYCLE = 360°
PULSE GENERATOR WAVEFORM (BEAM TURNED ON DURING INTERVAL ab≅3°)
+H.T.
R1
R2
C1
C2
V1
V2
TO QUENCH SOURCE FOR SYNCHRONIZATION (40 CYCLE MAINS SUPPLY)
Rg1
Rg2
OUTPUT
Rc
0.1μF, 2500V
200,000
0.1μF
-H.T.
PULSE GENERATOR
200,000
0.2
H.T.
-
+
CATHODE RAY OSCILLOGRAPH

Figure 3. Connection diagram for the pulse generator when connected to the cathode-ray oscillograph, the output wave form of the generator is also shown. R_1=100,000 ohms, R_2=100,000 ohms, R_c=2,500 ohms, R_{g1}=1.0 mega-ohm, R_{g2}=0.5 mega-ohm (for varying pulse width), C_1=0.05 microfarad, C_2=0.01 microfarad

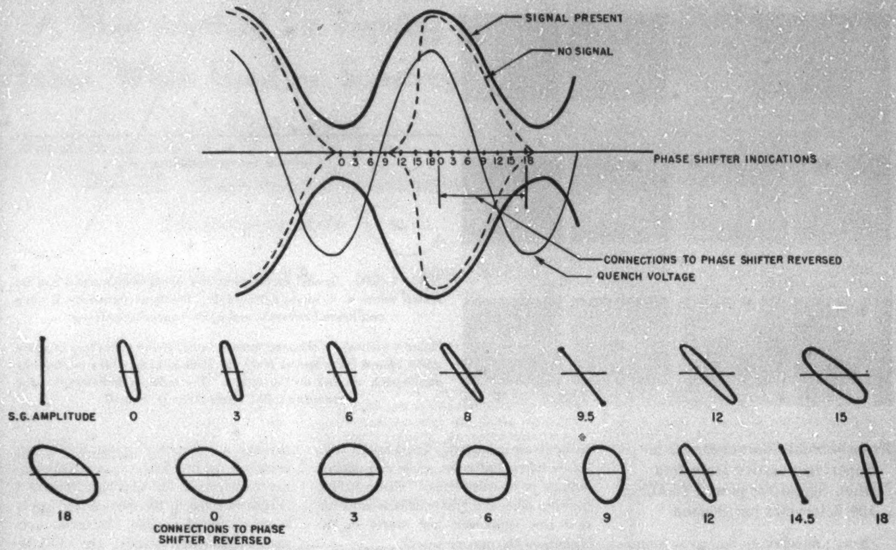

SIGNAL PRESENT
NO SIGNAL
0 3 6 9 12 15 18 0 3 6 9 12 15 18
PHASE SHIFTER INDICATIONS
CONNECTIONS TO PHASE SHIFTER REVERSED
QUENCH VOLTAGE
S.G. AMPLITUDE
0
3
6
8
9.5
12
15
18
0
CONNECTIONS TO PHASE SHIFTER REVERSED
3
6
9
12
14.5
18

Figure 5. Synchronization at signal frequency. Lissajous figures for samples of oscillation at various points along the envelope, high signal being applied. The numbers on the oscillograms give phase shifter indications. The amplitude of oscillation is indicated by a horizontal line. Signal frequency $f_1 \cong$ free oscillation frequency f_0. Enlarged oscillograms of the envelope of oscillation without signal and with signal are also shown

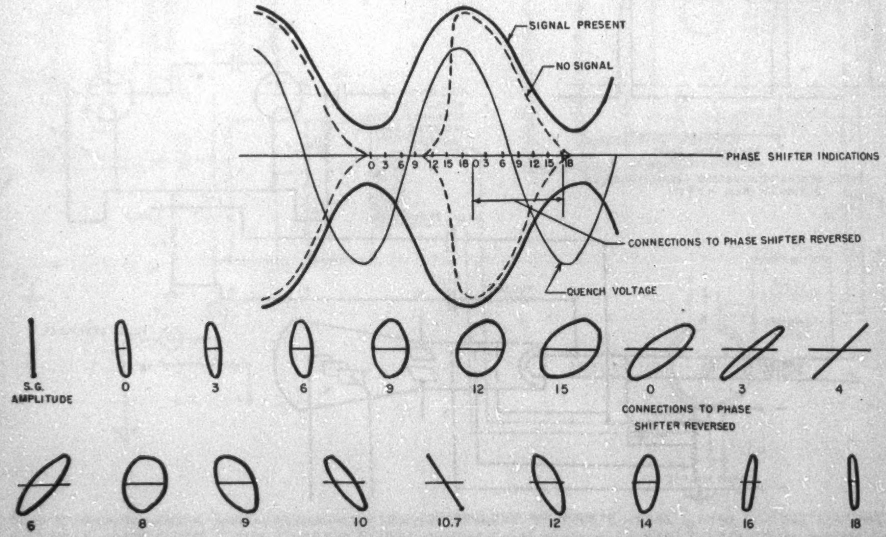
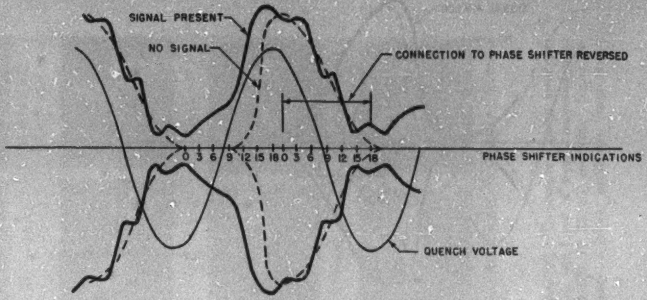

SIGNAL PRESENT
NO SIGNAL
0 3 6 9 12 15 18 0 3 6 9 12 15 18
PHASE SHIFTER INDICATIONS
CONNECTIONS TO PHASE SHIFTER REVERSED
QUENCH VOLTAGE
S.G. AMPLITUDE
0
3
6
9
12
15
0
3
4
CONNECTIONS TO PHASE SHIFTER REVERSED
6
8
9
10
10.7
12
14
16
18

Figure 6. Synchronization at the first side band. Lissajous figures as in Figure 5, but in this case signal frequency $f_1 \cong$ free oscillation frequency f_0 = quench frequency $p/2\pi$


SIGNAL PRESENT
NO SIGNAL
CONNECTION TO PHASE SHIFTER REVERSED
0 3 6 9 12 15 18 0 3 6 9 12 15 18
PHASE SHIFTER INDICATIONS
QUENCH VOLTAGE

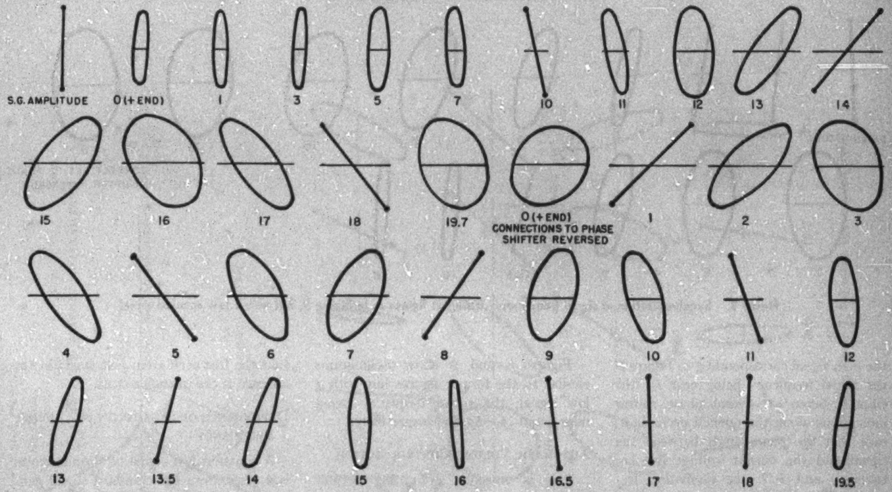

S.G. AMPLITUDE
0 (+ END)
1
3
5
7
10
11
12
13
14
15
16
17
18
19.7
0 (+ END) CONNECTIONS TO PHASE SHIFTER REVERSED
1
2
3
4
5
6
7
8
9
10
11
12
13
13.5
14
15
16
16.5
17
18
19.5

Figure 7. Synchronization at the fifth side band. Lissajous figures as in Figure 5, but $f_1 \cong f_0 - 5p/2\pi$

shows some irregularities. With a reasonable signal amplitude, the irregularities disappear and the envelopes were very steady. When the signal departs from the free oscillation frequency by approximately an integer number n of the quench frequency, the envelope of oscillation is modulated by the same number. The number n, for which the envelopes remain very steady, depends on the amplitude of the signal; as the signal increases, n increases. For higher values of n, the oscillations are no longer steady. The steadiness of the envelopes shows that the oscillations are synchronized during the whole quench cycle. Synchronization is either at the signal frequency f_1 in which case $n=0$, or at an equivalent side-band frequency $(f_1+np/2\pi)$. There is still a further test for such synchronization. Figure 2 shows several oscillograms at no signal and at comparatively high signal which is intentionally taken high in order to show the envelope modulation clearly. The signal frequency f_1 being near f_0 and $(f_0-np/2\pi)$, where f_0=free oscillation frequency, $p/2\pi$=quench frequency=40 cycles per second, and n=0, 1, 2, 3, 4, 5, 6, 8, 12, and 15. Steady envelopes can be obtained for as high n as 15, after which synchronization ceases.

Lissajous Figures for Samples of Oscillation at Various Instants in Quench Period, Synchronization at Signal Frequency and Equivalent Side Bands

To show the Lissajous figures corresponding to the oscillation at a certain instant during the quench period, apply the output of the detector to the X-plates and the signal generator output to an amplifier, then to the Y-plates of the cathode-ray oscillograph. A pulse generator is switched on in order that the electron beam appears only for a very short interval of time during the quench period corresponding to approximately 3 degrees of the quench cycle. Figure 3 shows a connection diagram for the pulse generator and its connection to the grid of the cathode-ray tube, as well as its output wave form. Figure 4 shows a calibrated quench cycle, showing clearly the indication of the phase shifter and the instant in the quench cycle which appears on the screen of the oscillograph.

Figure 5 shows several oscillograms at

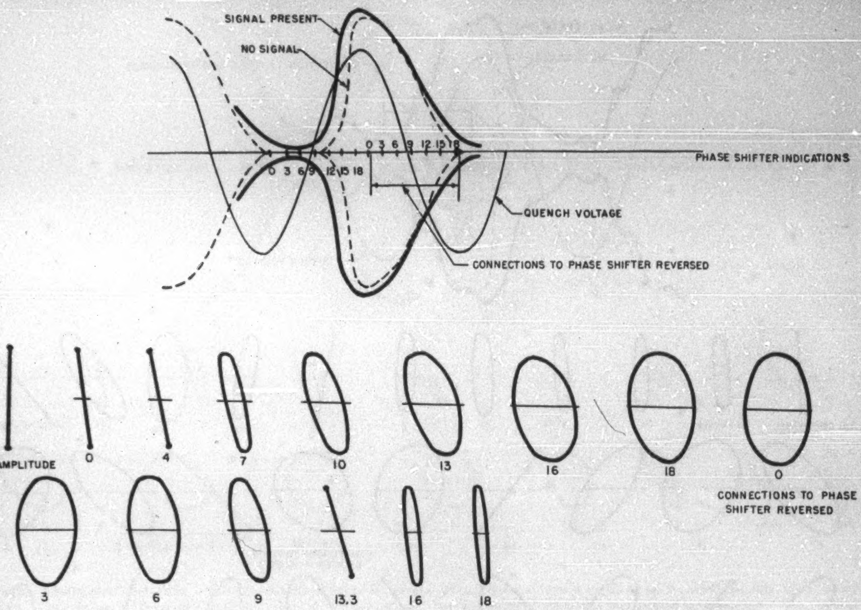


Figure 8. Synchronization at signal frequency. Lissajous figures as in Figure 5, but with a low applied signal

the high signal corresponding to Figure 2, the signal frequency being near f_0; the ellipses, taken at several phase shifter indications along the quench cycle, indicate that the phase angle between the signal and the output voltage lies between 0 and $\pi/2$, the oscillation frequency being that of the signal and there being synchronization at signal frequency. The corresponding amplitudes of oscillation are also shown.

Figure 6 shows several oscillograms at the same signal amplitude, but the signal frequency is near $(f_0-p/2\pi)$; the ellipses, taken at several phase shifter indications along the quench cycle, show that the phase angle between the signal and output voltage varies from 0 to $(\pi/2)+2\pi$, that is, repeats itself indicating that the frequency of oscillation $=f_1+p/2\pi$ and that there is synchronization at the side band number 1 which is termed semi-synchronization.

Figure 7 shows similar oscillograms at the same signal amplitude, but the signal frequency is near $(f_0-5p/2\pi)$. The oscillograms indicate that the frequency of oscillation is $(f_1+5p/2\pi)$ and there is synchronization at the side band number 5.

Figures 8 and 9 show oscillograms similar to the former figures but with a low signal, the signal frequency being near f_0 and $(f_0-5p/2\pi)$ respectively.

Automatic Volume Control Action

This is actually a limiting action. Such action can be understood from Figure 10 which represents several time variations of the envelopes at signals proportional to 0.0, 2.0, 5, 10, 20, 50, and 100, signal frequency being very near f_0. Figure 11 shows the increase of grid current and the decrease of plate current for increasing signal amplitude, signal frequency being very near f_0. It is clear from Figure 11 that the limiting action is a characteristic of a super-regenerative detector under logarithmic mode of operation.

High Frequency Detector Operating Frequency Around 1,000 Kilocycles per Second

Figure 12 represents a complete connection diagram, the values of the different elements are indicated. The d-c plate voltage is taken 100 volts. The coupling between plate and grid coils is so adjusted that the free oscillation just starts in the absence of the quench source.

Dependence of Sensitivity on Quench Frequency

A constant low signal of 50 microvolts is applied from the standard signal generator, signal frequency being near f_0 which is 1,100 kilocycles per second. A compensation for the grid current at no signal is effected by a 2-V battery and a 17,500-ohm resistance as shown.

Figure 13 shows the increase of grid current and the decrease of plate current with the variation of the quench frequency from 100 to 2,000 cycles per second, indicating clearly that there is an optimum quench frequency. The latter depends on the amplitude of the quench voltage V_q. For $V_q=60$ volts, $(p/2\pi)$ optimum$\cong$800 cycles per second, and for $V_q=50$ volts, $(p/2\pi)$ optimum$\cong$470 cycles per second.

Figure 14 shows several oscillograms representing the cartesian (time) envelopes at quench frequencies of 200, 500, 1,500, and 2,000 cycles per second, $V_q=60$ volts, no signal being applied. Figure 15 is as Figure 14 but with a constant applied signal. It is clear from

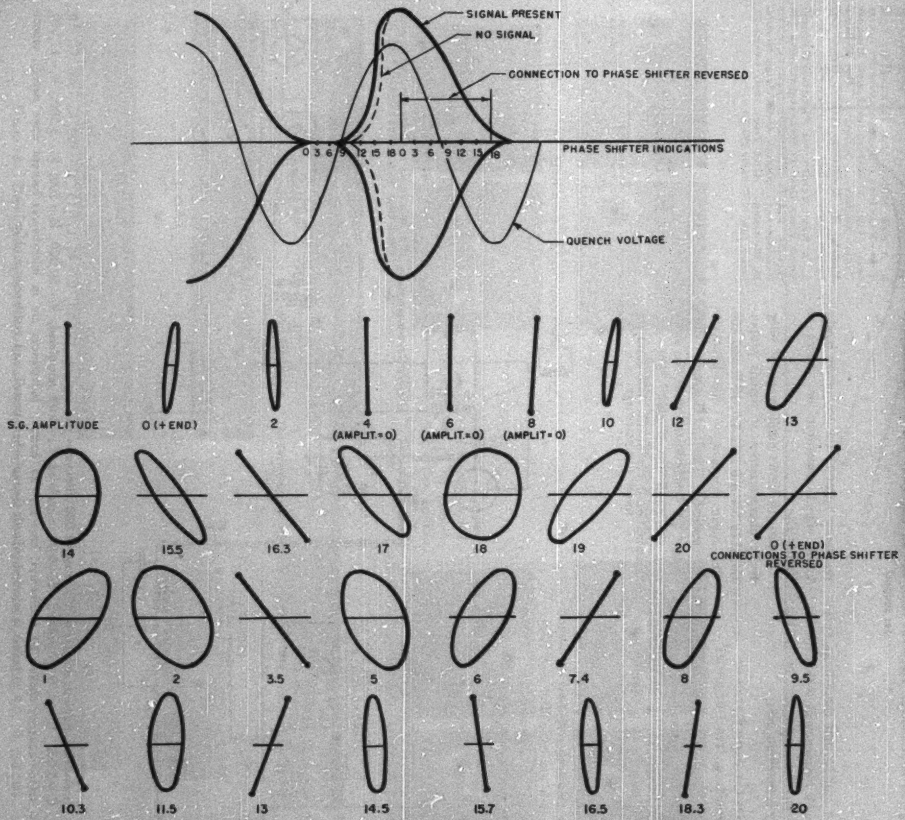

SIGNAL PRESENT
NO SIGNAL
CONNECTION TO PHASE SHIFTER REVERSED
PHASE SHIFTER INDICATIONS
QUENCH VOLTAGE
0 3 6 9 12 15 18 0 3 6 9 12 15 18
S.G. AMPLITUDE
0 (+END)
2
4 (AMPLIT.=0)
6 (AMPLIT.=0)
8 (AMPLIT.=0)
10
12
13
14
15.5
16.3
17
18
19
20
0 (+END)
CONNECTIONS TO PHASE SHIFTER REVERSED
1
2
3.5
5
6
7.4
8
9.5
10.3
11.5
13
14.5
15.7
16.5
18.3
20

Figure 9. Synchronization at the fifth side band, Lissajous figures as in Figure 7, but with a low applied signal

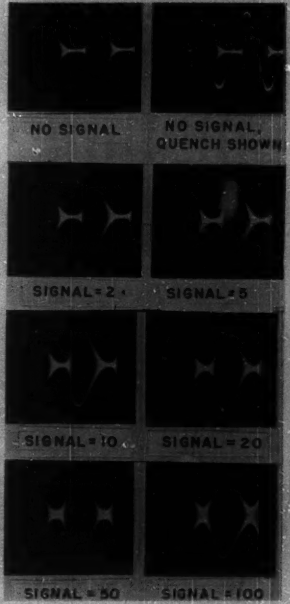

NO SIGNAL
NO SIGNAL, QUENCH SHOWN
SIGNAL = 2
SIGNAL = 5
SIGNAL = 10
SIGNAL = 20
SIGNAL = 50
SIGNAL = 100

Figure 10. Oscillograms giving envelopes at no signal and at signals proportional to 2.0, 5.0, 10, 20, 50, and 100. Signal frequency $f_1 \cong$ free oscillation frequency f_0

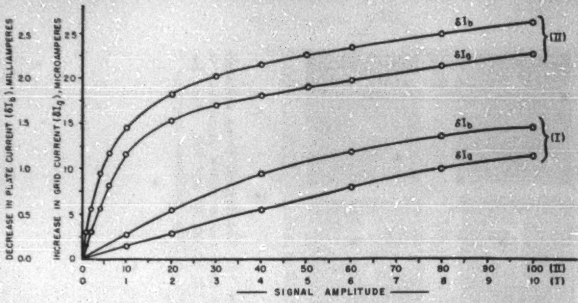


Figure 11. Automatic volume control action (limiting action) at no signal, $I_b=11$ milliamperes, $I_g=30$ microamperes, $V_{HT}=120$ volts, $V_q=40$ volts (rms), and $f_1=96$ kilocycles per second

Figures 14 and 15 that there is an optimum quench frequency at which the increase of envelope area and hence the sensitivity is maximum.

Selectivity Characteristics of Superregenerative Detectors

A static point-by-point selectivity measurement takes rather a long time and small drifts in the oscillation frequency occur. For this reason a dynamic characteristic is preferred. The signal is applied from a frequency modulated signal generator; the output voltage across the grid leak resistance is filtered in a 2-section low-pass filter and is applied to the vertical amplifier of the C.R.O. whose time base sweep drives the reactance valve of the frequency modulated signal generator as shown in Figure 12.

Figure 16 shows oscillograms of selectivity characteristics at signals proportional to 0.5, 1.0, 2.0, 4.0, 10.0, and 100. The mean frequency = 1,050 kilocycles per second. The oscillograms show clearly the synchronization at the signal and side-band frequencies. The synchronization curve at any side-band frequency is clearly indicated. At the high signal levels (10 and 100), the synchronizing curves of most of the frequency bands overlap and a smooth curve only is indicated. The oscillograms show a slight deformation near the ends; this is due to the transients of the oscillograph amplifier at the low sweep frequencies used. The synchronization curves at the frequency bands near the ends of the selectivity characteristics are not stable; after such ends, certain movable irregularities appear indicating noise characteristics.

(Figures 13 through 16 appear on the following page.)

References

1. A New Method for Treating Electron Tubes When Used as Superregenerative Detectors. Part I—Superregenerative Circuits Under No-Signal Condition, **Abd El-Samie Mostafa, M. El-Shishini.** *AIEE Transactions*, volume 72, part I, May 1953, pages 207–13.

2. A New Method for Treating Electron Tubes When Used as Superregenerative Detectors. Part II—Superregenerative Circuits Under Signal Condition, **Abd El-Samie Mostafa, M. El-Shishini.** *AIEE Transactions*, volume 72, part I, July 1953, pages 283–89.

No Discussion

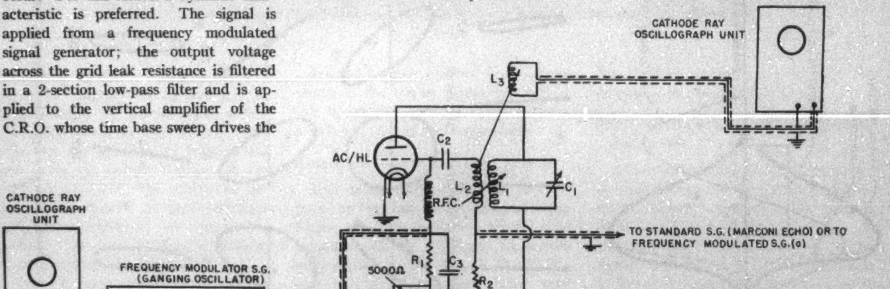

CATHODE RAY OSCILLOGRAPH UNIT
L_3
AC/HL
C_2
L_2
L_1
C_1
R.F.C.
TO STANDARD S.G. (MARCONI ECHO) OR TO FREQUENCY MODULATED S.G. (a)
R_1
C_3
R_2
5000Ω
M.A
R_4
COMPENSATING CIRCUIT
CATHODE RAY OSCILLOGRAPH UNIT
FREQUENCY MODULATOR S.G. (GANGING OSCILLATOR)
o
X
E
OUTPUT
f
R_5
A(Y)
E
X
(c)
OUTPUT
Ch_I
Ch_{II}
C_8
C_7
C_6
L.P.F. (2-SECTION π)
C_4
V_q
C_5
V_{HI}
M.A
R_3
− H.T. +
5000Ω
AMPLIFIER
f
10000
BEAT FREQUENCY AUDIO OSCILLATOR
MATCHING TRANSFORMER

Figure 12. Connection diagram, operating frequency = 1,000 kilocycles per second. $L_1=15$, $L_2=45$, $L_3=15$, $C_1=500$, $C_3=500$, $C_2=50$, $C_4=1{,}000$, $C_5=5{,}000$, $C_6=0.02$ microfarad, $C_7=0.04$ microfarad, $C_8=0.02$ microfarad, $R_1=50{,}000$, $R_2=500$, $R_3=5{,}500$, $R_4=17{,}500$, $R_5=50{,}000$, radio-frequency choke = 10 millihenrys, $ch_1=ch_2=25$ henrys. Inductances are in microhenrys (except when stated), capacitances in micromicrofarads (except when stated), and resistances in ohms

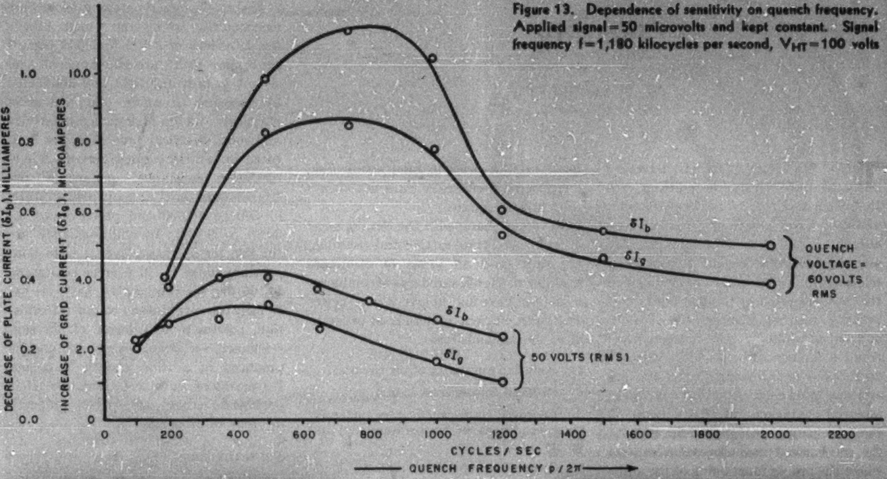


Figure 13. Dependence of sensitivity on quench frequency. Applied signal = 50 microvolts and kept constant. Signal frequency f = 1,180 kilocycles per second, V_{HT} = 100 volts

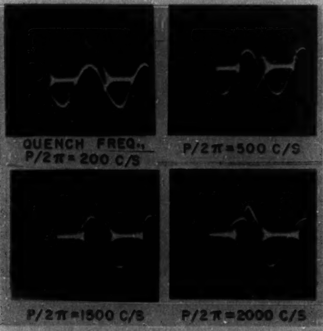
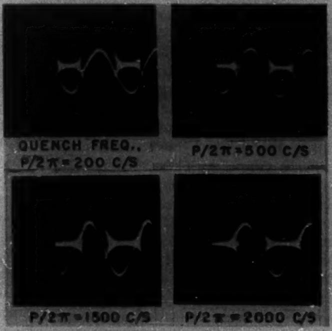

QUENCH FREQ., P/2π = 200 C/S
P/2π = 500 C/S
P/2π = 1500 C/S
P/2π = 2000 C/S

Figure 14 (left). Oscillograms giving envelopes without signal for various quench frequencies p/2π, quench voltage being also shown

Figure 15 (right). Oscillograms giving envelopes with a constant signal for various quench frequencies p/2π, quench voltage being also shown


QUENCH FREQ., P/2π = 200 C/S
P/2π = 500 C/S
P/2π = 1500 C/S
P/2π = 2000 C/S

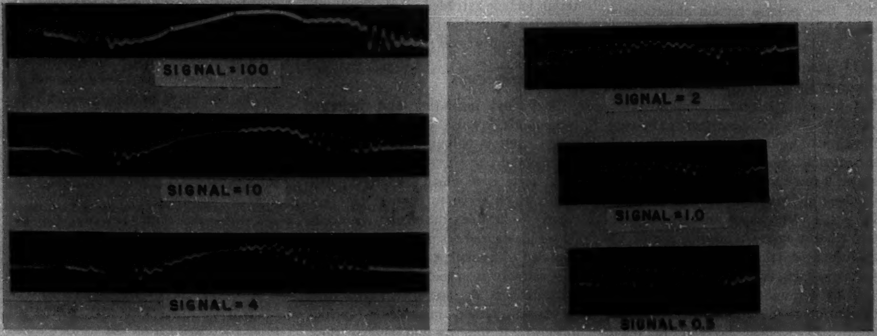

SIGNAL = 100
SIGNAL = 10
SIGNAL = 4
SIGNAL = 2
SIGNAL = 1.0
SIGNAL = 0.5

Figure 16. Oscillograms giving selectivity characteristics at signals proportional to 0.5, 1.0, 2.0, 4.0, 10, and 100

Wire and Cable in the Telegraph Industry

W. F. MARKLEY
NONMEMBER AIEE

THE TECHNOLOGICAL advances in the art of telegraphy, starting with the original simple Morse key, sounder, and relay, and progressing to the semiautomatic, and then the fully automatic selective-switching system in a modern telegraph office of today, have justified the modernization of the wire and cable network, insuring efficiency of operation and dependability of the complex electronic equipment, which is in continuous operation over a wide range of frequencies and operating conditions. An extremely essential component of this network is the wire plant concentrated in each of the mechanized communication centers where the proper functioning of the automatic equipment is dependent upon the performance of some 3,000 miles of high-grade wire conductors and more than 1,000,000 intricate wire connections.

General

Almost from the inception of the telegraph over a century ago, rubber compounds, enamel with textile overserves, and dry paper have comprised the insulating materials employed for inside and outside wires and cables. Deep-sea submarine cable has always relied upon gutta-percha as the medium of conductor insulation. Although all these materials have served the industry well, and even today are being employed to a considerable extent in the communication field, they possess certain objectionable characteristics.

Rubber compounds lack heat stability and resistance to deterioration in outdoor exposure, possess high electrical losses in high-frequency operation, and adversely affect weights and diameters. The combination of enamel coatings and textile servings is not adequate electrically under warm moist room conditions and involves high installation and terminating costs. Dry paper for cable conductors always necessitates the use of the traditional lead sheath or other impervious jacket to exclude moisture and is expensive to terminate and splice. Gutta-percha for deep-sea submarine cable has given an excellent account of itself, except that its uniformity of composition, heat stability prior to final installation, and electrical efficiency have not always been fully adequate.

In recent years, with the development of high-speed transmitting and recording apparatus and all forms of electronic equipment, the old standards of conductor insulation have had to give way to newer and more efficient materials in order to meet the demands for:

1. Increased sensitivity of the circuits;
2. Reduced transmission losses;
3. Improved electrical efficiency under adverse conditions of temperature and humidity;
4. Greater heat and light stability;
5. Reduced weights and diameters;
6. Decreased installation costs

These increased demands for more effective primary insulating materials have had a marked influence on the rapidity of development of synthetic-resin compounds, which began to appear about 15 years ago. From the very beginning, many of these so-called plastics found a ready market because of the many advantages they offered with respect to physical and electrical stability, chemical inertness, heat and light aging, moisture resistance, weight, cost, and so forth.

The plastic compounds employed for extruded primary insulation include vinyl compounds, polyethylene compounds, and fluorocarbon resins. The jacketing materials include extruded vinyl compounds, extruded polyethylene compounds, and extruded and woven nylons.

VINYL COMPOUNDS

The vinyl insulating and jacketing compounds consist of synthetic base resins, such as polyvinyl chloride or copolymers of polyvinyl chloride and polyvinyl acetate, to which are added plasticizers which serve to convert the rigid resins to flexible, rubberlike materials. These plasticized resins are further compounded with small percentages of certain base salts, such as lead salts, which serve as stabilizers to insure heat and light resistance. Other ingredients include lubricants to improve and increase gloss, fillers to provide opacity and decrease cost, and colorants to effect a limitless variety of shades.

These elastomeric compositions are mechanically strong and tough, having tensile values of 1,500 to 3,000 pounds per square inch; are nonflammable, resistant to heat and light and unaffected by exposure to water, oils, and most chemicals; exhibit high surface resistivity and good electrical properties, as indicated by a dielectric constant of 4.5 to 8, insulation resistance constant K per 1,000 feet of wire, at room temperature, of 500 to 10,000 and power factor of 0.070 to 0.160. In addition, they are suitable for operation over a wide temperature range (−50 degrees centigrade to 105 degrees centigrade), have an extremely low rate of water-vapor transmission, possess a high degree of abrasion resistance, are nonoxidizing, and can be produced in a wide variety of colors. By means of suitable formulation, it is possible to accent any specific property or group of properties.

POLYETHYLENE

Polyethylene, originally developed in England, is a pure hydrocarbon resin derived from the polymerization of ethylene gas. The principal advantages in the use of this material over vinyl compounds are its exceptionally low dielectric losses (dielectric constant 2.3 and power factor 0.0004) at all frequencies, its inherent extreme resistance to moisture, and its low temperature (−45 degrees centigrade flexibility). Polyethylene is reasonably tough and flexible as indicated by its tensile strength of 1,400 to 3,000 pounds per square inch and elongation of not less than 400 per cent at ordinary room temperature. It is extremely light in weight (specific gravity 0.92) and contains no plasticizers. Additives in small quantities (less than 2 per cent total, except as will be noted in reference to ocean cable insulation) comprise antioxidants to insure retention of electrical properties under thermal abuse, pigments to produce colored compounds, and carbon black to impart resistance to aging in outdoor exposure. The principal limitations to the use of polyethylene

---

Paper 53-164, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted October 7, 1952; made available for printing December 18, 1952.

W. F. MARKLEY is with the Western Union Telegraph Company, New York, N. Y.

The author acknowledges his indebtedness to the Technical Service Staff of Bakelite Company, Division of Union Carbide and Carbon Corporation, in proofreading and checking the technical data.

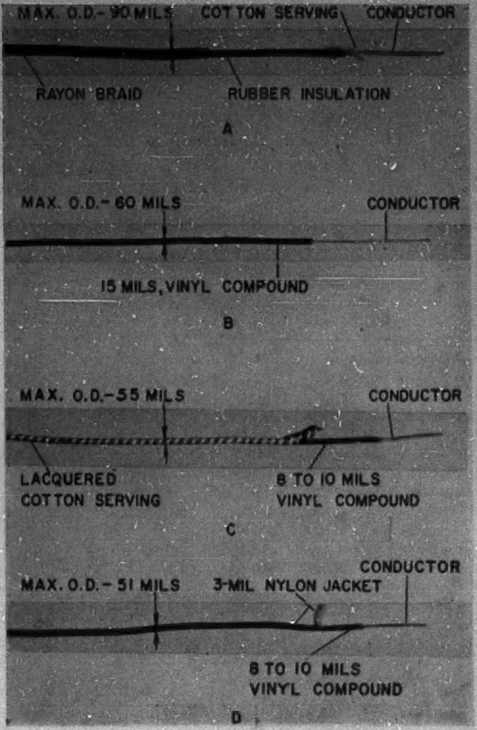


Figure 1 (left). Number 22 AWG instrument wire. Successive standards established during the past 12 years

A. Western Union standard—1937 to 1940
B. Western Union standard—1940 to 1946
C. Western Union standard—1946 to 1951
D. Western Union standard—1951 to present time

Figure 2 (above). 6-channel face plate, for multiplex distributor 226-A, illustrating the intricate wiring, forming, and terminating in this unit, requiring some 1,500 feet of number 22 AWG flexible vinyl-insulated and nylon-jacketed instrument wire. "A" shows one of the handmade forms, before applying the vinyl tubing indicated at "B"

in wire and cable insulations and jackets are thermal in nature; that is, temperatures above 100 degrees centigrade appreciably soften the material and, in the presence of an open flame, polyethylene will burn freely.

Fluorothene

In addition to the vinyl and polyethylene compounds limited use is made of extruded fluorothene synthetic resin as a primary insulation. Chemically, this resin is a polymer of chlorotrifluoroethylene. In its natural state this material is colorless, although it can be blended with solid fillers and coloring agents if desired. Fluorothene was selected as the insulating medium for the conductors used for wiring submarine cable repeaters, primarily because of its extreme toughness and abrasion resistance, and its resistance to the oil in which this equipment is immersed, as will be referred to later. It is likely that other fluorocarbon resins, such as teflon, would be equally satisfactory for this application. All these resins are characterized by extreme inertness to chemicals, unusual heat resistance, toughness over a wide range of operating temperatures, and good electrical properties. Manufacturing costs of the base resins and processing costs of the wire and cable coverings are relatively high, as compared with similar costs for vinyl and polyethylene compounds.

Nylon

Nylon, because of its abrasion resistance and extreme toughness and thermal stability over a wide temperature range, is employed primarily for jacketing purposes, although it was tried as a primary insulation in the extruded form, in connection with the development of office cable. E. I. du Pont de Nemours and Company describes nylon as a generic name applied to all materials defined scientifically as synthetic fiber-forming polymeric amides having protein-like chemical structure, derivable from coal, air, and water, or other substances. Nylon is supplied in various grades, the electrical grade being the one specified in the following as the protective covering for insulated wire. It is employed both in the extruded form as a continuous covering and in the form of a braid woven from nylon yarn.

Instrument Wires

Number 22 American Wire Gauge

The first major application of plastic-insulated wire in the telegraph industry occurred in 1940 when the supply of rubber was critical. Because millions of feet of number 22 American Wire Gauge (AWG) solid instrument wire are consumed annually as an all-purpose wire for general use in wiring equipment and apparatus, it seemed appropriate, to facilitate delivery of this badly needed material, to replace the rubber insulation with vinyl compound.

Prior to 1940, this instrument wire comprised a tinned-copper conductor insulated with a 15-mil wall of rubber in-

Figure 3. Rotary switch rack

sulation and an outer treated rayon braid, providing an over-all diameter of 90 mils, see Figure 1(A). After intensive study, the rubber and rayon covering was replaced with a 15-mil radial thickness of vinyl insulation, without any external jacket of any kind, and suitable for operating temperatures up to 60 degrees centigrade, Figure 1 (B). The over-all diameter of this wire was only 60 mils. This reduction of one-third in the diameter of the finished wire was a decided advantage in that it permitted a greater number of wires to be pulled into ducts, and effected decreased diameters of hand-made multiple conductor cable forms.

Instrument wire must normally withstand considerable abuse in handling and installation. It is bent around steel supporting strips and angles in cabinets and on frames, is held in multiple conductor forms by wrappings of twine, tape, and so forth is pulled across and is looped up against other wires causing considerable pressure at the contact surfaces. All these represent areas of potential weakness in the insulation due to thinning out of the compound due to cold flow, thereby resulting in excessive current leakage, or by completely cutting through to the conductor. Such wire failures and consequent outages of equipment became so numerous that for a time it looked as though plastic insulated wire could not be depended upon to give adequate service in automatic equipment in continuous operation.

At this point it became obvious that some form of mechanical reinforcement for the insulation must be resorted to if plastic compounds were to be continued for equipment wiring and, accordingly, in 1946 a lacquered-cotton serving was adopted as a jacket for this wire using number 100s single-ply cotton yarn. Because of the mechanical reinforcement by the cotton serving, the thickness of the vinyl insulation was safely reduced from the 15 mils specified to a maximum of 10 mils, Figure 1 (C).

This wire, with its lacquered-cotton serving, was generally satisfactory but still lacked the desired degree of thermal stability and abrasion resistance, showed objectionable wicking action in the presence of moisture, and possessed surface resistivity values under abnormal conditions of humidity that varied over a wide range after a period of service (primarily because of the tendency for dirt and foreign matter to collect in the interstices of the lightly-lacquered cotton serving).

In order to overcome these weaknesses the lacquered-cotton serving was replaced by an extruded electrical-grade nylon jacket in a 3-mil wall, and vinyl insulation heretofore suitable for 60-degree-centigrade operating temperature was changed to one suitable for 80 degree-centigrade operation. These revisions were adopted in 1951 and this wire is the present standard, Figure 1(D).

The outstanding advantages offered by the nylon jacket include the following:

1. Improved toughness (minimum 7,000 pounds per square inch) and abrasion resistance;

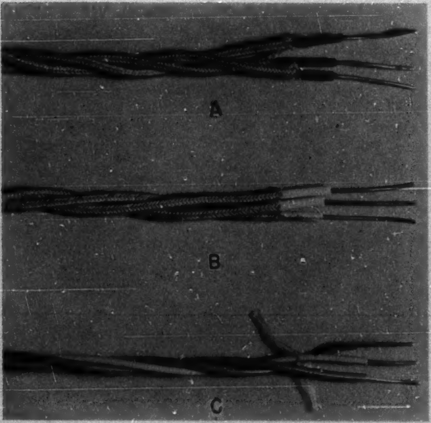


Figure 4. Triple-conductor office wire, number 20 AWG

A. Original standard-rubber insulation and cotton-braid covering; outside diameter, 107 mils (each conductor)
B. Intermediate standard-vinyl compound replaced rubber insulation
C. Present standard-nylon jacket substituted for cotton braid; outside diameter, 70 mils (each conductor)

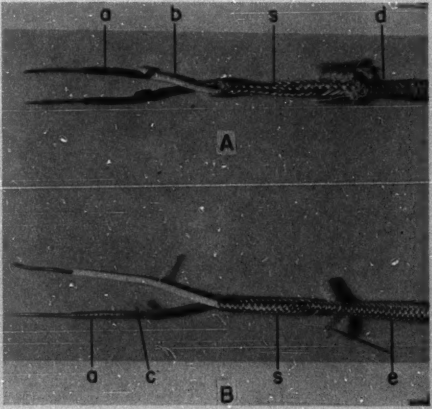


Figure 5. Shielded twisted pair wire, number 22 AWG

A. Old standard
B. Present standard
a = vinyl insulation
b = lacquered-cotton serving
c = nylon jacket
s = woven metallic shield
d = over-all lacquered-cotton braid
e = over-all extruded-nylon jacket

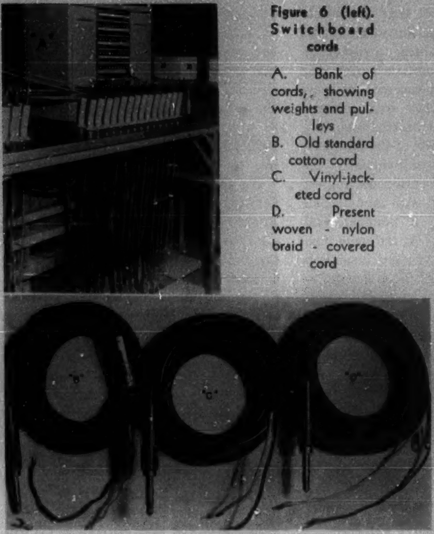


Figure 6 (left). Switchboard cords

A. Bank of cords, showing weights and pulleys
B. Old standard cotton cord
C. Vinyl-jacketed cord
D. Present woven - nylon braid - covered cord

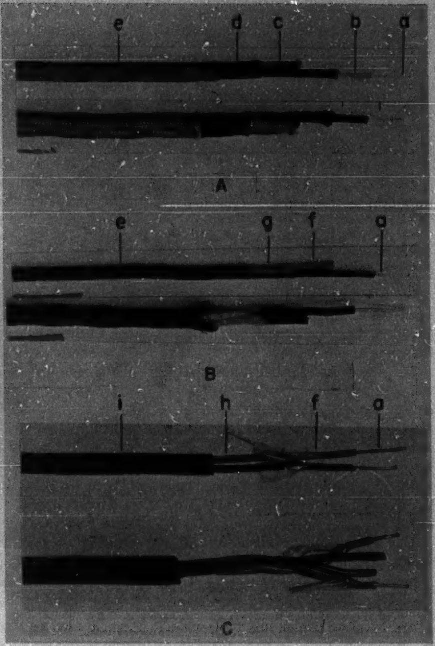


Figure 7. Ticker cables. 2-conductor size 2 number 20 AWG, parallel lay; 4-conductor size 4 number 18 AWG, spiral 4

A. Old standards
B. Intermediate standards
C. Present standards
a = tinned conductor
b = rubber insulation
c = cotton braid
d = rubber-filled tape serving
e = over-all cotton braid
f = vinyl insulation
g = paper-tape serving
h = nylon jacket
i = over-all vinyl jacket

2. Resistance to wetting, thereby effecting vastly improved surface resistivity;

3. Freedom from oxidation at temperatures close to 100 degrees centigrade;

4. Greater resistance to fungus;

5. Improved soldering and terminating characteristics;

6. Superior thermal stability;

7. Reduced loss of plasticizer from the primary insulation

This number 22 AWG solid vinyl-insulated nylon-jacketed wire, having an over-all diameter of 51 mils (0.051 inch), provided such outstanding performance as a trouble-free wiring medium that it has now been adopted for general use in making cross connections on main distributing frames, and for making connections on carrier racks and switchboards where heretofore 25-mil vinyl-insulated 22-gauge wire without a jacket (over-all diameter 75 mils) was employed, this heavy wall of insulation having been specified originally because of the cold-flow characteristics of the vinyl. This change effected a one-third reduction in over-all diameter. A companion wire consists of number 22 AWG flexible (stranded) instrument wire, employed for wiring printers and other apparatus having moving parts, as illustrated in Figure 2. The development of this 22-gauge instrument wire with its extruded thin walls of plastic insulation and jacket, required to meet severe electrical requirements, represents a novel design in the wire and cable industry.

NUMBER 24 AMERICAN WIRE GAUGE

Even more spectacular was the development of a number 24 AWG instrument wire, required primarily for wiring rotary switches. Figure 3 shows a multiplicity of these switches, assembled on a rotary switch rack. It is the author's opinion that the design of this wire represents one of the first major advances in the art of ultra-thin-wall extruded-vinyl insulation having superior electrical properties.

From the outset, this wire was required to meet unusual conditions with respect to over-all diameter, electrical performance, and ease of stripping of the insulation. The compact rotary switches on which this wire is employed are equipped with several hundred soldering lugs mounted in rows in close proximity. In connecting a number of these switches in series, this 24-gauge wire is required to pass between the lugs, a condition which accounts for the limitation on the over-all diameter of the finished wire. In addition, this wire must be capable of being stripped at intermediate points along the wire, when making attachment to corresponding lugs on successive switches.

It was indicated that the thickness of primary insulation would have to be held to a maximum of 6 mils because of the space limitation. To provide this extremely thin wall of insulation, pressure-sensitive self-sealing insulating plastic tapes, which appeared to be the only practical method available, were tried but found inadequate.

An attempt was then made to extrude a good electrical vinyl compound on the conductor in a nominal thickness of 5 mils.

This request presented a real problem to the manufacturer for the following reasons:

1. Difficulty of obtaining tinned-copper wire having adequate smoothness;

2. Extent of delicate handling required during extrusion process to prevent frequent conductor breakage;

3. Adequate centering of conductor in the thin wall of compound;

4. The question of insulating and processing technique that would insure a minimum insulation resistance of 200 megohms per 1,000 feet, at 15.5 degrees centigrade, for full spools of finished wire after a water immersion period of 6 hours.

After a number of trials, a 5-mil vinyl insulated 24-gauge wire was economically produced that conformed to all of the foregoing requirements.

To provide mechanical protection for the primary insulation, a lacquered-cotton serving was specified, similar to that already described for 22-gauge instrument wire, experience with nylon having been extremely limited at that time. The nominal over-all diameter of this finished wire was only 36 mils (0.036 inch).

This wire gave an excellent account of itself in service, except that complaints were received that the lint or fluff from the cotton serving was being carried off by air currents in air-conditioned operating rooms and lodging in the moving arms of the rotary switches with adverse operating effects. In order to obviate this trouble, and in view of the increased knowledge of nylon subsequently acquired, the lacquered-cotton serving on this wire was replaced with an extruded nylon jacket having a nominal wall thickness of $2^1/_2$ mils. At the same time the thickness of the primary insulation was increased to 7 mils (from 5 mils) to comply with the request of several manufacturers to effect easier processing. With this change the insulation resistance was increased to 500 megohms (instead of 200 megohms) per 1,000 feet when immersed in water at 15.5 degrees centigrade for 6 hours. The nominal over-all diameter of this finished wire, the present standard, measures about 40 mils (0.040 inch).

NUMBER 28 AMERICAN WIRE GAUGE

Even though tremendous advances have been made in the development of electronic devices, the relay continues to play an important role in the operation of modern telegraph systems. In fact, the relay can perform many functions better and more economically than the electron tube. In certain polar relays the lead wires for the cores are subjected to abnormal temperatures and severe mechanical abuse in installation and service.

For this application, number 28 AWG flexible copper wire, insulated with a 10-mil wall of fluorothene, was originally specified in order to provide adequate thermal stability to withstand the high ambient temperature of about 100 degrees centigrade (prevailing during continuous operation) and to insure adequate mechanical stability, in view of the bending encountered near terminating points. Subsequently, at a greatly reduced cost for the wire, the fluorothene insulation was replaced with vinyl compound, suitable for 105-degree-centigrade operation, over which was applied a 3-mil wall of extruded nylon. There would have been some apprehension about the use of the vinyl-insulated conductor for this service, but the reinforcement, from both a thermal and a mechanical stand-

Figure 8 (below). Western Union Office cables

A. Old Western Union standard; enamel insulation and textile servings
B. Plastic design, polyethylene insulation, and lacquered-cotton serving
C. Present all-plastic standard, vinyl insulation and nylon

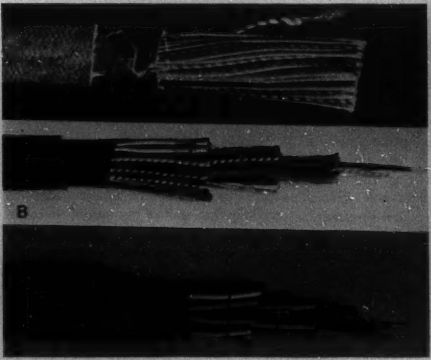
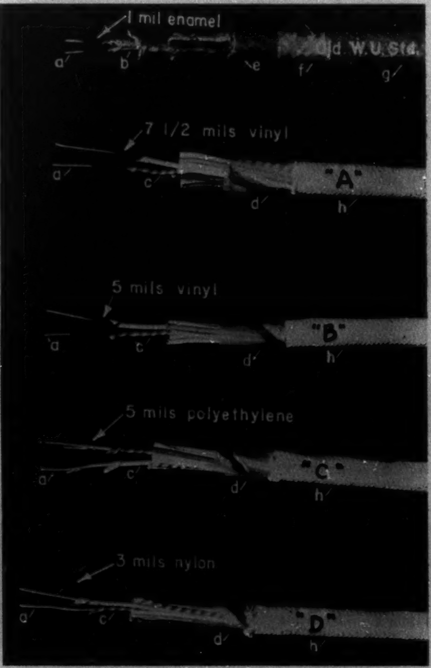


Figure 9 (right). Experimental office cables

a = tinned conductor
b = rayon and cotton servings
c = lacquered-cotton serving
d = polyethylene-core tape
e = paper-tape servings
f = cotton-ribbon serving
g = heavy painted-cotton braid
h = lacquered-cotton braid



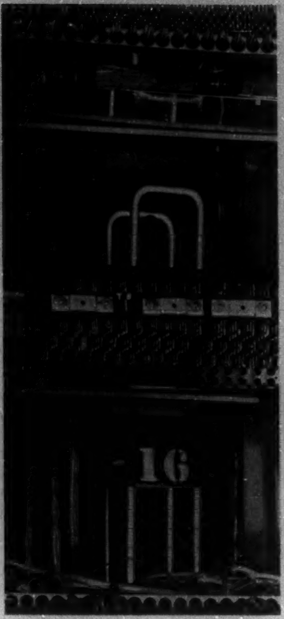


Figure 10. Section of transmitter finder, showing installation of 40-pair polyethylene-insulated office cable

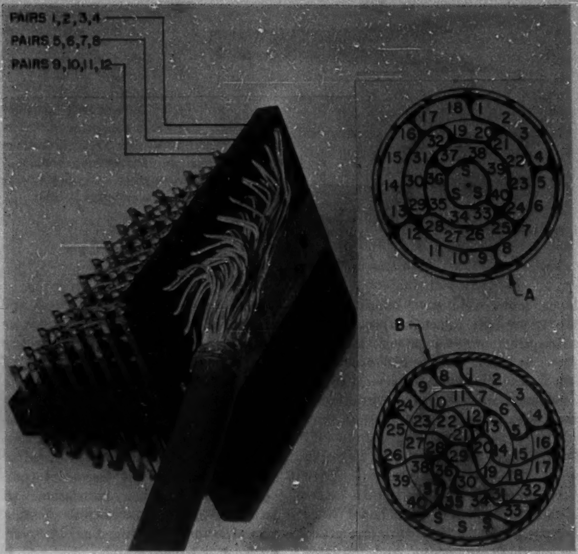


Figure 11. Termination of butt of 40-pair (plus 3 spare pairs) polyethylene insulated office cable, on standard Western Union terminal block, indicating the position of successive groups of 4 pairs at each 8-conductor level

A. Old standard method of positioning the pairs
B. New method adopted for the all-plastic cable

point, by the nylon jacket has provided a finished wire having adequate serviceability. Cable forms, comprising 70 of these wires, have been satisfactorily employed for wiring test tables.

OTHER TYPES

In addition to the various types of instrument wires referred to in the foregoing, limited use is made of other sizes of solid and stranded conductors, insulated with vinyl compounds and an outer jacket of extruded nylon, such as the station or office wire illustrated in Figure 4. The nylon jacket is always held to an extremely thin wall because of the toughness of the nylon and its adverse effect on the stripping characteristics of the covering.

Experience has indicated that polyethylene insulation, in the extremely thin walls that are specified for instrument wire, is much more fragile and susceptible to mechanical injury during processing than vinyl compounds. Consequently, polyethylene is used for instrument wire only in exceptional cases where uniformly low capacitance is required over a wide range of frequencies and under extreme variations in temperature and humidity.

All colors for coding purposes consist of solid colors embodied in the primary insulation using light and dark tints of various shades to effect multiplicity of colors desired. All of these colors are sharp and readily distinguishable from one another through the clear nylon jacket. Consideration has been given to employing colored stripes on the primary insulation for coding purposes, but experience has indicated that with the various inks used for this purpose there is a serious adverse effect of the ink solvent on the insulation resistance of the primary insulation. One manufacturer has been reasonably successful in applying colored stripes to the nylon jacket but this method of coding has not been satisfactory for the extremely small over-all diameters of the finished instrument wires generally employed in the telegraph plant.

Shielded Twisted Pair Wire

Considerable time was devoted to the development of a number 22 AWG twisted pair wire, having an over-all woven-metallic shield, required for certain circuits in electronic equipment. Initially, in the original design, shown in Figure 5(A), there was some question as to how much of the surface of the twisted pair wire should be covered by the woven metallic braid.

For test purposes to determine this feature, lengths of twisted pair wire were made up having three nominal degrees of shielding, namely 70-, 85-, and 100-per-cent coverage. The results of the electrical tests indicated that no significant difference, in crosstalk, attributable to the various degrees of shield coverage, could be detected. The major contribution of the shield appeared to be the minimizing of direct capacitance between circuits. From a mechanical standpoint, that is, flexibility and minimum damage to the shield during severe bending and abuse in installation, the shield having the least coverage was the most desirable.

While, at the low frequencies used on these pairs, it appeared that a coverage of 30 to 50 per cent would be sufficient, it was decided, in view of the fact that commercial shields usually represented about

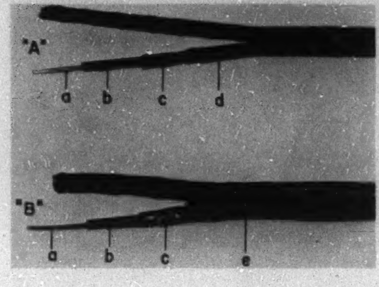


Figure 12. Outside distributing wires

A. City distribution wire
B. Tree wire
a = untinned conductor
b = 1/32-inch vinyl insulation
c = Kraft paper tape
d = number 30s 2-ply weatherproof cotton braid
e = number 10 Seine-twine weatherproof braid

70-per-cent coverage, to standardize on the 70-per-cent value for this wire. A subsequent improvement in this wire, shown in Figure 5(B), consisted of replacing the lacquered-cotton serving on the conductors and the treated-cotton outer braid with extruded nylon.

Switchboard Cords

Figure 6 illustrates the various stages in the development of switchboard cords. For many years these cords comprised copper tinsel conductors, each insulated with servings of cellulose-acetate yarn, treated with a moistureproof compound over which was applied a soft cotton braid. These conductors, together with suitable fillers, were laid up to form a core which was covered with several cotton braids, the outer braid consisting of hard glazed cotton, see Figure 6(B). These cords are subjected to considerable wear and abuse in service, as shown by Figure 6(A), because of the continuous flexing and abrasion of pulleys and weights

Because the hardest wear on these cords occurs in the outer braid, it was decided to substitute for the all-cotton cords, a plastic cord, shown in Figure 6(C) wherein extra-flexible stranded conductors were employed, each insulated with a thin wall of extruded vinyl compound, laid up into a core with suitable fillers, and covered with an over-all extruded-vinyl jacket. These cords, however, proved unsatisfactory, because of the difficulty of obtaining a thermoplastic jacket compound that possessed adequate flexibility and abrasion resistance without the undesirable characteristic of having the plasticizer ooze out from the compound during warm weather, a condition which made the cords sticky and extremely difficult to handle.

In order to overcome the undesirable features of the vinyl-covered cords, the plastic jacket was replaced by a woven nylon braid, see Figure 6(D), at which time a bronze tinsel conductor was adopted to provide superior strength and flexibility characteristics. Indications are that the service life of these cords will be several times that of the ordinary glazed cotton braided cords.

Ticker Cable

Practically all of the cables required for inside use have been redesigned to provide for plastic primary insulation and jacketing compounds. Figure 7 is a typical example showing the old, intermediate, and present styles of 2- and 4-conductor ticker cables. The final designs provide exceptional ease of installations and excellent performance in service, particularly in warm moist climates.

Office Cable

Perhaps the most important development in inside cable is concerned with the redesign of the enamel and textile-insulated multipair cable, employed in large quantities for wiring offices. This cable is supplied in sizes varying from 3 pair to 50 pair. For a great many years the standard for this type of cable consisted of 22 AWG solid-copper conductors, each enamel-insulated and covered with servings of silk and cotton yarn, which were laid up in pairs, with a concentric lay, into a core that was thoroughly saturated with a natural or mineral beeswax, and paraffin mixture. Over this core were applied several spiral wrappings of heavy paper tape to serve primarily as a barrier for the flameproofing paint that was heavily applied to the over-all woven cotton braid which served as the jacket. This old design is shown in Figure 8(A).

This cable is subjected to considerable abrasion and abuse during installation in that it is pulled through conduit, run up over steel bars and angles of supporting frames, given sharp bends, is tied in place on racks in groups, and so forth. Also, much of this cable is employed in relatively short lengths, involving numerous terminations. Frequently, long lengths of the jacket and paper tape must be removed to provide sufficient length for breaking out the conductors at terminating points. The principal disadvantages in the use of this cable for many years have been the relatively high labor cost of terminating and the adverse effect of moisture on the operation of the circuits. In damp and wet locations, it was always necessary to supply this cable with a lead sheath. With increased knowledge on the performance and efficiency of plastic compounds, it became evident that a vast improvement could be made in the design of this cable.

POLYETHYLENE CABLE

About 5 years ago, four special plastic cables were made up in the 7-pair concentric-lay size, each having a different type of extruded plastic insulation with a lacquered-cotton overserve on the conductors; over the core was applied a 10-mil polyethylene tape and a cable jacket consisting of a lacquered woven cotton braid. For comparative test, a standard enamel and textile cable in 5-pair size was obtained.

Figure 9 shows the appearance of these four plastic cables, and the old Western Union standard enamel and textile-insulated cable. In the *A*, *B*, *C*, and *D* cables in this figure, the lacquered-cotton serving is all about the same general character and thickness; in the standard enamel-insulated cable, this yarn covering comprises two rayon servings and a serving of marking cotton, all of which have a certain amount of mineral beeswax impregnation.

Insulation resistance and capacitance tests were conducted on 100-foot lengths of these five cables after exposure for various periods of time to different at-

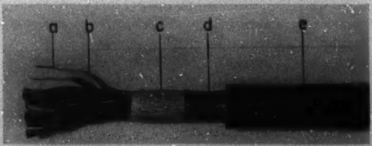


Figure 13. Plastic cable for outside use to replace paper-insulated lead-covered cable

a = untinned conductor
b = 20-mil polyethylene insulation
c = crepe-paper bedding
d = 5-mil copper tape
e = over-all black-polyethylene jacket

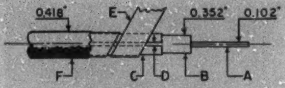


Figure 14. Carrier cable

A. Number 10 AWG untinned hard-drawn copperweld wire (40-per-cent conductivity; tensile strength, 1,130 pounds; d-c resistance, 2.55 ohms per 1,000 feet)

B. Polyethylene insulation (brown-bakelite DE-6490)

C. Five-mil soft untinned copper tape, applied longitudinally, equivalent to number $11^{1}/_{2}$ AWG

D. The maximum gap shall be 1/16 inch, and the edges of the copper tape may touch occasionally

E. Three-mil soft untinned copper tape, spirally applied without overlap; nominal gap between adjacent convolutions is 1/16 inch

F. Number 30 2-ply soft-cotton closely woven braid (maximum braid angle = 62.5 degrees), weatherproofed according to Western Union Specification 1638-C

Weight of Component Parts of Cable

Conductor, 10 gauge	= 0.029 pound per foot
Polyethylene, 1/8-inch wall	= 0.034 pound per foot
Copper tape, 5 mils	= 0.020 pound per foot
Copper tape, 3 mils	= 0.012 pound per foot
Finished cotton braid	= 0.012 pound per foot
Total	= 0.107 pound per foot = 107 pounds per 1,000 feet

mospheric conditions varying from a temperature of 7.2 degrees centigrade (45 degrees Fahrenheit) and 35-per-cent relative humidity, to 35 degrees centigrade (95 degrees Fahrenheit) and 90-per-cent relative humidity. In addition, voltage-breakdown tests were made on conductors taken from the cable, and on short lengths of the cable that had been subjected to severe flexing and compression, to simulate conditions encountered in service. These tests definitely indicated the outstanding superiority of the polyethylene, and the poor insulating value of nylon and enamel. The $7^{1}/_{2}$-mil vinyl-insulated cable was somewhat superior to the 5-mil vinyl-insulated cable.

While both the $7^{1}/_{2}$-mil vinyl and the 5-mil polyethylene-insulated cables overcame the insulation weakness of the old standard enamel and textile-insulated cable, the fact that the polyethylene gave such outstanding electrical performance prompted the decision to give it a field trial. Accordingly, limited quantities in 40- and 50-pair size cables were made up, with a nominal 5-mil wall of polyethylene insulation and lacquered-cotton serving, with a vinyl tape placed longitudinally over the cable core and a vinyl jacket employed for the outer sheath, as illustrated in Figure 8 (B). These cables were installed in several reperforator offices as illustrated in Figure 10. The ease of installation and termination were favorably commented upon by the installers and the performance of these cables to date has been extremely satisfactory under all weather conditions.

VINYL CABLE

During the manufacture of these polyethylene-insulated cables, it was observed that the thin wall of insulation was extremely fragile and easily damaged during processing and cabling. Inasmuch as the superior electrical characteristics of polyethylene were not required for this service, and in view of the greater toughness of thin vinyl compounds, and furthermore since the $7^{1}/_{2}$-mil vinyl-insulated cable gave adequate electrical performance in the tests on the original experimental plastic cables previously referred to, it was decided to substitute an 8- to 10-mil vinyl compound and a 3-mil nylon jacket for the polyethylene insulation, and lacquered-cotton serving respectively. In other words, the standard 22-gauge instrument wire, previously referred to, was adopted for these cable conductors, an arrangement which simplified production schedules and factory inspection. This design, illustrated in Figure 8 (C), represents the present standard. This same cable is now employed in damp and wet locations where formerly lead-covered cable was required.

A further refinement in the design of the plastic cable consisted of providing a lay-up for the pairs that permitted the conductors to break out easily from their natural position in the cable during the terminating operation, thus reducing to a minimum the bending and twisting of the conductors at the soldered butts. Figure 11 (A) shows the positioning of the conductors in the old 40-pair enamel-insulated cable, wherein the pairs follow in numerical sequence around the cable in the order of termination. In Figure 11 (B), the pairs of the plastic cable follow in numerical sequence back and forth across the cable. From the heavy irregular outlines in the diagrams it can be observed, when mounting the cable butt on a standard terminal block, that severe distortion of the conductors takes place as successive groups of four pairs break-out in the enamel cable, whereas with the plastic cable these groups fan out in their normal position.

RELATIVE COSTS

The cost of the all-plastic cable in the smaller sizes is somewhat greater than that of the old enamel-insulated cable but in the larger sizes it is appreciably cheaper. Also, the time consumed in terminating the plastic-insulated cable, due to the ease of stripping, is only 50 to 60 per cent of that required for the enamel type, resulting in a substantial labor saving, inasmuch as a large percentage of this cable is installed in short lengths varying from 8 to 30 feet, thereby calling for the termination of from 6 to 24 butts per 100 feet of cable. In addition, with the plastic cable, all of the stripping on a number of lengths can be done at one time and all of the soldering work carried out at a later time regardless of the interval between stripping and soldering. In the enamel type of cable, it is necessary to solder almost immediately after stripping because the copper conductors, bared during the stripping operation, readily tarnish and oxidize.

ADVANTAGES OF ALL-PLASTIC CABLE

The principal advantages in the use of the all-plastic cable over the former standard, having enamel-insulated conductors, can be summarized as follows:

1. Improved moisture resistance;
2. Superior electrical and physical properties and flexibility;
3. Increased resistance to fungus;
4. Better flame resistance;
5. Sharper coding colors for the insulated conductors;
6. Greater ease of installation and terminating, effecting vastly reduced installation costs.

Outside Wires and Cables

DISTRIBUTING WIRE

Single and twisted pair insulated and weatherproof braid-covered, tinned number 16 AWG hard-drawn copper and copperweld wire, commonly referred to as outside distributing wire, is employed in city distribution work in the telegraph plant to the extent of several thousand miles per year, in providing and maintaining time, call, ticker, printer, and other services between offices and subscribers, and to provide drops from open-line wire into railroad stations, signal towers, and industrial plants. In addition, this wire is used in repair and restoration service in storm areas, in reconstruction work involving pole-line changes necessitated by reason of grade-crossing eliminations, highway improvements, and so forth.

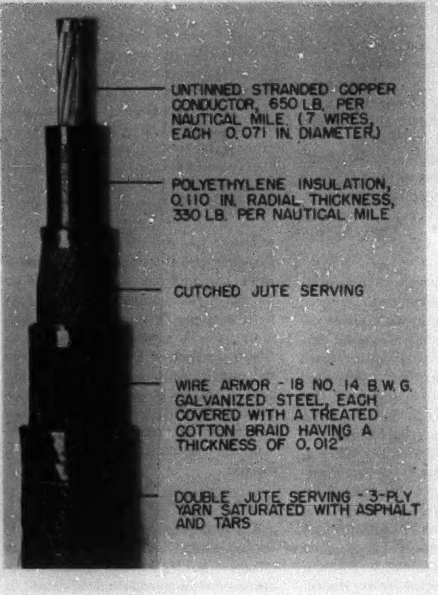

UNTINNED STRANDED COPPER CONDUCTOR, 650 LB. PER NAUTICAL MILE. (7 WIRES, EACH 0.071 IN. DIAMETER.)
POLYETHYLENE INSULATION, 0.110 IN. RADIAL THICKNESS, 330 LB. PER NAUTICAL MILE
CUTCHED JUTE SERVING
WIRE ARMOR - 18 NO. 14 B.W.G. GALVANIZED STEEL, EACH COVERED WITH A TREATED COTTON BRAID HAVING A THICKNESS OF 0.012"
DOUBLE JUTE SERVING - 3-PLY YARN SATURATED WITH ASPHALT AND TARS

Figure 15 (left). Deep-sea submarine cable, single conductor, type D

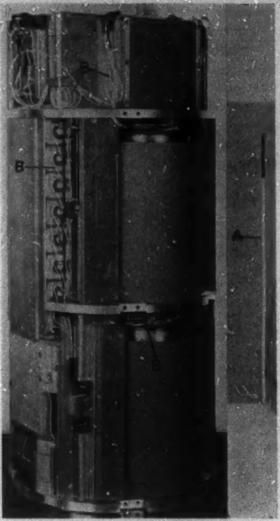


Figure 16 (right). Submerged telegraph repeater with outer steel casing removed

A shows number 18 AWG solid silver-plated copper conductor, insulated with 10 mils of fluorothene, employed for the firing at B

The first sections at which the covering of this wire begins to fail are where the wire is tied or clamped to supports, especially at points of abnormal strain, such as corners or dead ends; improper methods of attaching insulated wire at such points may reduce its useful life by 50 per cent or more. More severe abuse to the wire occurs in reconstruction work and in the restoration of service in storm areas when this wire is pulled over fences and cross-arms and under rails, tied to poles, trees, and so forth, and allowed to remain in water and ice for long periods.

Although the original rubber insulation on this wire gave a reasonably good account of itself in this service for many years, it possessed several weaknesses which not only shortened the life of the wire, but resulted in its inadequacy for certain types of service. In 1942, Western Union pioneered in the development of the first self-supporting thermoplastic-insulated wire, designed for this type of outdoor service in the communications field. In the early consideration of plastic insulation for this application, there was considerable apprehension about the suitability of this type of material, in view of the rough handling during installation and service, and the fact that the finished wire is subjected to an extremely wide range of weather conditions throughout the United States.

After considerable laboratory work, a vinyl compound was selected for this service. As an external covering, the application of a weatherproof braid appeared to be the most advantageous type of mechanical protection, from the standpoint of both economy and performance, that could be satisfactorily adapted for use over the insulated conductor. In the first complete wires that were prepared, it was found that the saturating and finishing compounds employed for weatherproofing the braid adversely affected the vinyl compound. The volatile component of the hot saturating compounds degraded the electrical characteristics of the vinyl insulation, and the heat of application of these compounds softened the insulation to the extent that the conductor would float in the insulation during processing of the braid. To overcome these weaknesses a thin Kraft paper tape was applied longitudinally over the insulation and simultaneously with the braid. Figure 12 shows the design of two types of this wire. The estimated life is in excess of 25 years and the mechanical and electrical stability are superior to the previous rubber-insulated types.

For this service, consideration is being given to conductors insulated with an extruded 1/32-inch wall of so-called high-molecular-weight "rigid"-vinyl compound, modified slightly to improve flexibility. Rigid resins have sufficient plasticity to enable them to be extruded, worked, and processed without the addition of plasticizers. These resins have the outstanding electrical, physical, and chemical properties of the pure unmodified vinyl-base resins previously referred to. No outer jacket of any kind is applied to this wire, as the rigid vinyl appears to have adequate hardness and abrasion resistance over a wide temperature range. While the proper balance between hardness, flexibility, and impact resistance for this compound has not been worked out definitely, indications are, from tests conducted so far, that an adequate simple insulated wire of this type for outside service will be available in the near future.

LINE CABLE

For cable in city distribution service and pole-line runs, where the length is not too great, it is customary to employ lead-covered cable, comprising paired number 19 AWG dry-paper-insulated annealed-copper conductors laid up with a concentric lay. This cable, commonly referred to as *CNB* cable, is supplied in sizes varying from 6 pairs to 455 pairs, and is employed in quantity, both aerially and underground. It is relatively heavy and expensive, involves high installation and terminating costs, and in many congested localities necessitates high maintenance cost in aerial installations.

It is believed that the objectionable features of this cable can be overcome by the use of plastic-insulated and jacketed cable employing polyethylene insulation for the conductors, to provide the uniformly low capacitance required. In a nominal wall thickness of about 20 mils, polyethylene gives approximately the same range of mutual-capacitance values between the conductors of a pair as that obtained in the case of the standard paper-insulated cable. Indications are that this 20-mil wall of polyethylene will provide excellent transmission characteristics for circuits up to 50 kc.

Figure 13 shows the general design of a cable of this type that is being considered, in a 26-pair size, for an aerial installation on Long Island. This cable weighs only about half as much as an equivalent lead covered cable.

CARRIER CABLE

The demand arose for an inexpensive 2-conductor twisted pair, or simple cable, which would be suitable for short-haul (up to 100 miles) carrier communication at frequencies up to 30 kc per second and which would be self-supporting, to enable it to be strung on existing pole lines without the use of a messenger and, thereby, effect minimum installation costs.

At the outset, it was evident that hard-drawn copper or copper-covered steel would be required for the conductors to provide the desired mechanical efficiency, and that polyethylene would be necessary for the primary insulation to insure the low-transmission losses for this service. As the outer jacket for mechanical protection, experience has indicated that a modern weatherproof woven-cotton braid would be both adequate and economical.

Accordingly, in this initial investigation, twisted pair copper wires were considered, in both number 10 and number 13 AWG, each conductor insulated with polyethylene, and covered with an aluminum-faced paper tape and a weatherproof braid. Subsequent samples were provided with a single 4-mil metallic shield and weatherproof braid over the two insulated conductors as a unit. These designs showed prohibitive attenuation of 3 to 4 decibels per mile, necessitating the installation of repeaters every 7 or 8 miles.

Theoretical transmission studies were then made on a simplified coaxial-cable design comprising a single number 10 AWG copper-covered steel conductor insulated with a heavy wall of polyethylene, over which a second conductor comprising a soft 5-mil longitudinal copper tape (equivalent to number $11^{1}/_{2}$ AWG copper) is applied; for mechanical reasons a 3-mil spiral tape is applied over the longitudinal tape. The details of this construction are illustrated in Figure 14.

From calculations based on the dimensions in Figure 14, the ratio of inside diameter of the outer conductor to the diameter of the central conductor approximates 3.5 as compared with the ideal ratio of 3.6 to insure minimum attenuation for this design. Indications are that, with the outer conductor applied longitudinally so as to approximate a continuous solid tube, instead of a helical tape serving, a more effective design is obtained from the standpoint of electrical losses, particularly at television frequencies.

This cable design is considered mechanically and economically promising, and is believed suitable for television frequencies if required. The attenuation is estimated at 2.0 decibels per mile at 30 kc per second, which is high, but not entirely prohibitive. It is generally recognized that a gaseous dielectric would effect a marked reduction (ratio of 3 to 2) in the attenuation, but at present such a design seems economically impractical.

The experimental approach to this project, made possible by the procurement of a length of this cable for test purposes, now on order, is valuable, since the magnetic properties of the copper-covered steel core in this application are not too well known. Furthermore, the mechanical characteristics of the thin copper outer conductor employed in a self-supporting cable are also somewhat speculative. From tests that will be conducted on this experimental length, it should be possible to estimate with fair accuracy the properties of any moderate departure in the proposed design.

Submarine Cable and Equipment

Prior to 1940, gutta-percha was used almost universally as the insulating material for ocean cable conductors, rubber compounds having been found generally unsatisfactory. All gutta-percha cables had been made abroad, either in England or Germany. As World War II shut off the supply of gutta-percha even more effectively than rubber, pressure for the development of a suitable synthetic insulation for ocean cable increased.

Deproteinized rubber had been used in this country with some success and is still being employed because it offers equally good electrical properties and certain improved physical characteristics, as compared with gutta-percha. An even better solution was achieved by English scientists in adapting polyethylene (or "polythene," as it is called in England). In Europe, polyethylene compounds containing 10 to 15 per cent "Vistanex" (polyisobutylene) are also employed, because of the greater ease of processing with existing equipment, as compared with pure polyethylene. Since 1943, polyethylene-insulated ocean cable (see Figure 15) has been manufactured in the United States.

The first purchases of polyethylene cable by Western Union were made in 1944. At the present time, the Western Union Telegraph Company has in service approximately 800 miles of this type of cable, manufactured both here and abroad, and has on order about 600 additional miles.

Polyethylene compounds are more stable than gutta-percha at relatively high atmospheric temperatures, such as those encountered by a cable ship in southern waters, where the cable might be subjected to the direct rays of the sun for short periods before laying. Also, indications are that polyethylene compounds may be substantially resistant to attack by the teredo worm, whereas gutta-percha must be protected with a serving of copper or brass tape.

In Europe, polyethylene-insulated cable costs less than the gutta-percha type and, due to the present low dollar value of the pound sterling, is appreciably cheaper than any ocean cable produced in the United States. For cables manufactured in this country, the deproteinized-rubber type is cheaper than polyethylene-insulated cable.

Considerable data have been published by the Western Union Telegraph Company during the past 2 years covering the design, installation, and operation of intermediate submerged telegraph repeaters in a long transoceanic cable, which represents the first venture into a previously unexplored field and which permits tripling signaling speed in some cases.

Equipment housed in this repeater must be designed to insure reliable, unattended, long-life operation under unusual conditions. The instrument wire for this unit must be provided with an insulating compound having extreme toughness and abrasion resistance, to withstand the abuse during installation and service resulting from the sharp bending and flexing encountered, and must have superior chemical resistance against the action of the low-vapor-pressure oil in which the unit is immersed. Fluorothene insulation was specified for this application, as indicated in Figure 16 illustrating wiring technique.

Conclusions

The foregoing descriptions serve to illustrate the major uses of the newer insulating materials that have been adopted for various specific applications in the telegraph industry during the past 12 years in supplying the millions of conductor feet of insulated wire and cable installed annually. No attempt has been made to discuss all of the special services where these insulating materials have been specified with extremely satisfactory results.

In the old days, it was necessary to more or less adopt some standard wire or cable for a specific application, for which it was not always fully adequate. Today, it is possible to have the bare conductor, the insulating materials, and the finished cable tailor-made so as to insure efficient and lasting performance in service under almost any prescribed operating conditions. There still remain many applications where the suitability of the newer insulating materials is under intensive study.

New formulations and techniques, effecting reduced diameters and weights and superior physical, chemical, thermal, and electrical characteristics, are continually being developed, that forecast major advances of great significance in the art of insulated-wire and cable design and present an ever-widening future in this field, not only in the telegraph sphere, but in all American industry as well. It would be extremely conservative to predict that even greater advances in the development of more effective wire and cable insulating compounds will be brought to light during the next decade.

◆

Discussion

A. L. Duna (United States Steel Company, Worcester, Mass.): Mr. Markley's paper is an excellent portrayal of wire and cable problems which existed in the telegraph industry, and the changes in wire and cable design which eliminated many of the problems and materially reduced costs of installation. However, the new cable designs have created many problems for the cable manufacturer, and in essence it may be said that the problems of installation have been transferred to the manufacturer of the cable. This, of course, is as it should be. It is the way of progress—the responsibility of the manufacturers to make things better for the customer who is always right.

Let us consider just one of the relatively simpler problems which have confronted the cable manufacturer in the making of instrument wire and office cable. Just picture the small number 22 or 24 AWG solid copper wire, on which 8 to 10 mils of polyvinylchloride (PVC) resin must be extruded. That gives us a nominal thickness of 9 mils, with a tolerance of ± 1 mil, which must be extruded on a thread of wire traveling at high speed, and must be maintained continuously for millions of feet of production. Also, picture the 3- to 4-mil nylon jacket, which is extruded over the PVC insulation, with a tolerance of only ± 1/2 mil, also at high speeds, for continuous runs of millions of feet. The problem has now become one of precision manufacture where extreme vigilance and care in the operation is essential, for the mere dropping of an eyelid may result in the loss of thousands of feet of production, because of failure to meet the requirements of the specifications.

Pursuing the manufacturing problem further, the making of multiple-pair office cable, which is composed of instrument wires as already described, involves the handling of these relatively fragile wires in many other operations and equipments and the transportation of them on long journeys between operations. Unless extreme care is taken in handling the component parts of the cable as it progresses through to completion, damage becomes inevitable and results in test failures, short lengths, and eventual remakes of components which are beyond repair.

Notwithstanding our difficulties, we wish to commend Mr. Markley and Western Union for their foresight and courage in pioneering this radically new design of cable which, in spite of the manufacturing problems, is a forward step in the progress of wire and cable manufacture.

W. F. Markley: The author is grateful to A. L. Duna for his excellent presentation of the manufacturer's problems that develop when attempting to fabricate insulated wire and cable according to Western Union specifications.

It is quite true, as Mr. Duna points out, in referring to inside multipair cables, that the "relatively fragile wires" must withstand the handling encountered in many operations, involving long journeys between fabricating points, and that considerable care must be exercised in processing the insulated conductors to insure freedom from undue degradation of the physical and electrical characteristics of the insulated conductors. However, it has been the author's experience that after a manufacturer has developed the technique and systematized the production of these wire and cable products with the desired precision, extremely little trouble is experienced.

These insulated wires and cables are employed for wiring complex electronic equipment concentrated in highly mechanized communication centers throughout the United States, and operating continuously 24 hours a day over a wide range of frequencies and operating conditions. This fact makes it essential that extreme vigilance and care be taken in processing the products to provide the desired quality to insure adequate functioning of this automatic equipment. Years of experience of Western Union engineers have definitely indicated that installation practices and performance requirements justify the quality levels that have been standardized.

Effects of Harmonics on the Frequency of Oscillation as Well as on the Asymmetry of the Resonance Curves

ABD EL-SAMIE MOSTAFA
NONMEMBER AIEE

Synopsis: Transforming the nonlinear equations representing the free and synchronized oscillations to linear ones with periodically varying coefficients, the complete solutions of oscillator equations to the first power of the tube parameters are obtained. The relation between frequency variation and harmonic contents is given. The effect of the harmonics on the asymmetry of the resonance curves of a synchronized oscillator is obtained for the first time. Experimental verification has been carried out, and the results agree well with the theoretical predictions.

THE DEPENDENCE of the frequency of oscillation of an electron tube oscillator on the nonlinear properties of the maintaining circuit has been the subject of several investigations, theoretical and practical.[1,2] The effects of the harmonics on the asymmetry of the resonance curves of synchronized oscillators cannot be regarded as absolutely final and complete.[3] As far as the author knows, no previous investigator has obtained a clear relationship between the harmonic contents and the asymmetry of these resonance characteristics.

Expressing the nonlinear relationship of the electron tube by a power series of infinite number of terms, the complete solution of oscillator equations has been obtained. The harmonic contents for the free and synchronized oscillations are given. The effect of harmonic contents on the asymmetry of the resonance characteristics is rigidly obtained.

Theoretical Analysis

Free Oscillations

The differential equation for the output voltage v of the oscillator in Figure 1 is given by

$$d^2v/dt^2+(1/C)(G+di_p/dv)dv/dt+\omega_0^2v=0 \qquad (1)$$

where i_p is the a-c component of the plate current and $\omega_0^2=1/LC$. The losses in the circuit are represented by a shunt conductance G in order to simplify the analysis, and the results obtained can apply satisfactorily as well to the more usual case where the losses are associated with the inductance, especially when the circuit losses are relatively low. Such procedure of assuming a constant conductance G is deduced from the fact that the impedance of the parallel C-L-G circuit, for the fundamental as well as for the harmonic voltages, is very nearly the same as the more usual circuit. Strictly speaking, G cannot be considered as absolutely constant, and the analysis of the more usual case where the losses are associated with the inductance will be given by the author in another paper.

i_p can be represented as a power series of v, thus

$$i_p=\sum_{n=1} a_nv^n$$

giving

$$di_p/dv=\sum_{n=1} na_nv^{n-1}$$

The upper limit of the summation depends on the number of terms required in the expansion for satisfactory representation, but generally it is infinity.

For an oscillatory solution, v can be assumed to be of the form

$$V_t\cos\omega t+\sum_{m=2}^{\infty} V_{t_m}\cos(m\omega t+\theta_m)$$

where V_t, being a very slow function of time compared with the frequency of oscillation, is the fundamental amplitude at any instant t; V_{t_m} and θ_m are also slow functions of time and represent the amplitudes and phases of the harmonics respectively, and $\omega/2\pi$ is the frequency of oscillation (fundamental). For the solution of equation 1 to the first power (degree) of the tube parameters a_n, it is sufficient to substitute $V_t\cos\omega t$ for v in di_p/dv. Substitution gives

$$d^2v/dt^2+(G_t/C)\left(1+\sum_{r=1}^{n-1} h_r\cos r\omega t\right)\times dv/dt+\omega_0^2v=0 \qquad (2)$$

where

$$G_t=G+h_0 \qquad (3)$$

$$h_0=\sum_{n=1}[na_n/2^{n-1}\{(n-1)/2\}!]\times \{(n-1)(n-2)(n-3)\ldots \text{ up to } (n-1)/2 \text{ factors}\}V_t^{n-1} \qquad (4)$$
$$(n \text{ being odd})$$

$$G_th_r=\sum_{\substack{p=(r+1),\\(r+3),(r+5)\ldots}}[pa_p/2^{p-2}\{(p-r-1)/2\}!]\times \{(p-1)(p-2)(p-3)\ldots \text{ up to } (p-r-1)/2 \text{ factors}\}V_t^{p-1} \qquad (5)$$

In this way, the nonlinear differential equation 1 has been transformed to a linear one with periodically varying coefficients.

For the solution of equation 2 to the first power of a_n and hence to the first power of h_0 and G_th_r, the methods of superposition and successive approximation are adopted and given in Appendix I. The output voltage v will be given by

$$v=A_s\epsilon^{(1/2C)\int G_t\{(h_2/2)-1\}dt}\times \left[\cos\omega t-(1/2\omega C)\sum_{r=2}\{G_t/(r^2-1)\}\times (h_{r-1}-h_{r+1})\sin r\omega t\right] \qquad (6)$$

where A_s is a constant depending on initial conditions and G_th_r is to be obtained from equation 5 and $\omega\simeq\omega_0$.

It is clear from equation 6 that the instantaneous amplitude V_t builds up from the small starting value A_s to the stationary value V according to the relation

$$V_t=A_s\epsilon^{(1/2C)\int G_t\{(h_2/2)-1\}dt}$$

Stationary state is attained when $h_2=2$. Substitutions from equations 3, 4, and 5 give

$$\alpha_1-G=\sum_{p=3}[pa_p/2^{p-1}\{(p-1)/2\}!]\times \{(p-1)(p-2)(p-3)\ldots \text{ up to } (p-3)/2 \text{ factors}\}V^{p-1} \qquad (7)$$
$$(p \text{ being odd})$$

$$=(3/4)a_3V^2+(5/8)a_5V^4+(35/64)a_7V^6+(63/128)a_9V^8+(231/512)a_{11}V^{10}+\ldots \qquad (7A)$$

where $\alpha_1=-a_1=(\mu M/L-1)(1/\rho)$; ρ and μ are the plate resistance and the amplification factor of the tube respectively. Equation 7 gives the stationary possible fundamental amplitudes V of the output voltage.

Paper 53-159, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Winter General Meeting, New York, N. Y., January 19–23, 1953. Manuscript submitted September 10, 1951; made available for printing December 3, 1952.

Abd El-Samie Mostafa is with the Alexandria University, Alexandria, Egypt.

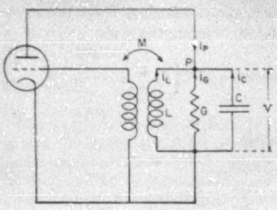


Figure 1. Tuned plate oscillator, polarized potentials omitted

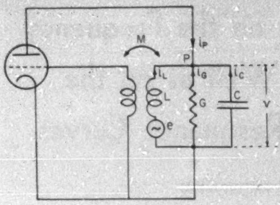


Figure 2. Simple synchronized oscillator, polarized potentials omitted

At the stationary state, equation 6 reduces to

$$v = V\left[\cos\omega t - (1/2\omega C)\sum_{r=2}\{G_1/(r^2-1)\}\times (h_{r-1}-h_{r+1})\sin r\omega t\right] \quad (8)$$

$$= V\cos\omega t - \sum_{r=2} V_r \sin r\omega t \quad (8A)$$

where

V_r = amplitude of the rth harmonic

$$= (V/2\omega C)\{G_1/(r^2-1)\}(h_{r-1}-h_{r+1}) \quad (9)$$

$G_1 h_r$ is the value of $G_t h_r$ obtained from equation 5 after substituting V for V_t. Therefore

$$V_r = (1/2\omega C)\{1/(r^2-1)\}\times \left[(ra_r/2^{r-2})V^r + \sum_{\substack{p=(r+2),(r+4),\\(r+6)\ldots}} \times [rpa_p/2^{p-2}\{(p-r)/2\}!]\{(p-1)(p-2)\ldots \text{ up to } (p-r-2)/2 \text{ factors}\}V^p\right] \quad (9A)$$

Putting $r=2, 3, 4\ldots$ the different harmonics can be obtained, see Appendix II.

It is clear from the previous analysis that to the first power of the tube parameters a_n, the frequency of oscillation $\omega/2\pi$ should equal to $\omega_0/2\pi$. If there is any deviation from $\omega_0/2\pi$, it must be of the second power of a_n. Approximations up to the second power can be obtained by substituting $\left(V_t\cos\omega t - \sum_{m=2} V_{tm}\times \sin m\omega t\right)$ for v in di_p/dv of equation 1 and following the same analysis but using closer approximations. However, the frequency deviation can be obtained adopting the following conception.

Due to the presence of the harmonics, intermodulation with each other and with the fundamental give additional terms in the fundamental plate current i_p which are not in phase with the output voltage v. A deviation of frequency from $\omega_0/2\pi$ is therefore necessary to restore the correct phase relation between current and voltage.

Substituting $\left(V\cos\omega t - \sum_{r=2} V_r\sin r\omega t\right)$ for v in i_p gives

$$i_p = \sum_{n=1} a_n v^n = \sum_{n=1} a_n\left(V\cos\omega t - \sum_{r=2} V_r \sin r\omega t\right)^n \quad (10)$$

Analyzing equation 10 to the first power of V_r and using equations 5 and 7 we get

$$i_p = -GV\cos\omega t - (1/2)\sum_{r=2} G_1(h_{r-1}-h_{r+1})\times V_r\sin\omega t + \text{components of frequencies other than } \omega/2\pi \quad (11)$$

Substituting equation 9 in equation 11 gives

Fundamental component of (i_p)

$$= i_{p\,\text{Fund.}} = -GV\cos\omega t - \omega C\times \sum_{r=2}(r^2-1)(V_r^2/V^2)\sin\omega t = I_{p\,\text{Fund.}}\cos(\omega t+\theta)$$

where

$$\tan\theta = (\omega C/-G)\sum_{r=2}(r^2-1)(V_r^2/V^2)$$

But, from C-L-G parallel circuit, Figure 1

$$\tan\theta = (\omega C/G)\{1-\omega_0^2/\omega^2\}$$

Therefore

$$(\omega_0^2/\omega^2)-1 = \sum_{r=2}(r^2-1)(V_r^2/V^2) \quad (12)$$

This is the Groszkowski[1] relation.

Equation 12 indicates that the frequency of oscillation is reduced due to the presence of the harmonics. Therefore, solution of equation 1 to a better order of approximation, concerning the frequency of oscillation, will be as follows

$$v = A_s \cdot \epsilon^{(1/2C)\int G_t\{(h_2/2)-1\}dt}\times \left[\cos\left(\int\omega_t dt\right) - (1/2\omega C)\sum_{r=2}\times \{G_t/(r^2-1)\}(h_{r-1}-h_{r+1})\times \sin\left(\int r\omega_t dt\right)\right] \quad (6A)$$

where the instantaneous frequency of oscillation $\omega_t/2\pi$ is given by

$$\omega_t^2/\omega^2 = 1 + \sum_{r=2}(r^2-1)(V_{t_r}^2/V_t^2) \quad (12A)$$

V_t and V_{t_r} are the instantaneous amplitudes of the fundamental and rth harmonic respectively. It is clear from equation 6(A) and 12(A) that as the amplitude of oscillation builds up, the frequency of oscillation decreases till the stationary state is reached and V_t and ω_t attain the stationary values V and ω given by equations 7 and 12 respectively.

FORCED OSCILLATIONS

The differential equation for the output voltage v of the synchronized oscillator of Figure 2 is given by

$$d^2v/dt^2 + (1/C)\{G + di_p/dv\}dv/dt + \omega_0^2(v-\epsilon) = 0 \quad (13)$$

where ϵ is an external source of amplitude E and frequency $\omega/2\pi$; G, i_p, v, and ω_0 have the same indications as before.

In this case, i_p can also be represented as a power series of v, thus

$$i_p = \sum_{n=1} a_n v^n, \text{ and } di_p/dv = \sum_{n=1} n a_n v^{n-1}$$

If synchronized oscillation only is present, whether the system is in the stationary or the transient state, the output voltage v will generally be of the form

$$v = b_t\cos\omega t + \sum_{m=2} b_{tm}\cos(m\omega t+\theta_m)$$

and the synchronizing voltage of the form, $\epsilon = E\cos(\omega t+\phi_t)$, where b_t, b_{tm}, θ_m, and the instantaneous phase angle ϕ_t are slowly varying functions of time compared with the frequency of oscillation. b_t is the fundamental amplitude and b_{tm} and θ_m are the instantaneous amplitudes and phases of the harmonics of the output voltage respectively.

For the solution of equation 13 to the first power of the tube parameters a_n, it is sufficient to substitute $b_t\cos\omega t$ for v in di_p/dt and to substitute $k_t\{v\cos\phi_t + (1/\omega)\sin\phi_t\, dv/dt\}$ for ϵ, where $k_t = E/b_t$. Substitution gives

$$d^2v/dt^2 + (G_t/C)\left(1+\sum_{r=1}^{n-1} h_r\cos r\omega t\right)\times dv/dt + \omega_{0t}^2 v = 0 \quad (14)$$

where

$$\omega_{0t}^2 = \omega_0^2(1-k_t\cos\phi_t) \quad (15)$$

$$G_t = G + h_0 \quad (3A)$$

$$h_0 = -\omega_0 k_t\sin\phi_t + \sum_{n=1}[na_n/\{2^{n-1}((n-1)/2)!\}]\{(n-1)(n-2)(n-3)\ldots \text{ up to } (n-1)/2 \text{ factors}\}\times b_t^{n-1} \quad (4A)$$

(n being odd)

$$G_t h_r = \sum_{\substack{p=(r+1),(r+3),\\(r+5)\ldots}} [pa_p/2^{p-2}\{(p-r-1)/2\}!]\{(p-1)(p-2)(p-3)\ldots \text{ up to } (p-r-1)/2 \text{ factors}\} b_t^{p-1} \quad \textbf{(5A)}$$

Equation 14 is of the same form as equation 2 or equation 23 and therefore its solution, being of the form $b_t \cos \omega_t + \ldots$ as assumed, can be directly obtained from equation 6(A) and is given by

$$v = A_s e^{(1/2C)\int G_1\{(h_2/2)-1\}dt} \times \left\{\cos \omega t - (1/2\omega C)\sum_{r=2} \{G_1/(r^2-1)\} \times (h_{r-1}-h_{r+1}) \sin r\omega t\right\} \quad \textbf{(16)}$$

where A_s is a constant depending on initial conditions and $G_t h_r$ is to be obtained from equation 5(A), and $(\omega_{0t}^2/\omega^2) = 1+\sum_{r=2} (r^2-1)b_{t_r}^2/b_t^2$; using equation 15 we get

$$-k_t \cos \phi_t = \{2(\omega-\omega_0)/\omega_0\} + \sum_{r=2} (r^2-1)b_{t_r}^2/b_t^2 \quad \textbf{(17)}$$

It is clear from equation 16 that the instantaneous amplitude b_t of the synchronized oscillation builds up from the small starting value A_s according to the following relation

$$b_t = A_s e^{(1/2C)\int G_1\{(h_2/2)-1\}dt}$$

Stationary state is attained when $h_2=2$, in which case $b_t=b$, $b_{t_r}=b_r$ and $\phi_t=\phi$; substituting from equation 3(A), 4(A), and 5(A), and using equation 7 we get

$$\omega_0(E/b)\sin\phi = -(\alpha_1 - G) \times \left[1 - \frac{\displaystyle\sum_{p=3,5,7\ldots} [pa_p/2^{p-1}\{(p-1)/2\}!] \times \{(p-1)(p-2)\ldots \text{ up to } (p-3)/2 \text{ factors}\} b^{p-1}}{\displaystyle\sum_{p=3,5,7\ldots} [pa_p/2^{p-1}\{(p-1)/2\}!] \times \{(p-1)(p-2)\ldots \text{ up to } (p-3)/2 \text{ factors}\} b_0^{p-1}}\right] \quad \textbf{(18)}$$

where b_0 is the amplitude of the free oscillation, and $\alpha_1 = -a_1$.

Also, at the steady state equation 17 becomes

$$-\omega_0(E/b)\cos\phi = z + \omega_0 \sum_{r=2} (r^2-1)b_r^2/b^2 \quad \textbf{(17A)}$$

where

$$z = 2(\omega - \omega_0)$$

Elimination of ϕ between equations 18 and 17(A) gives the frequency/amplitude equation, namely

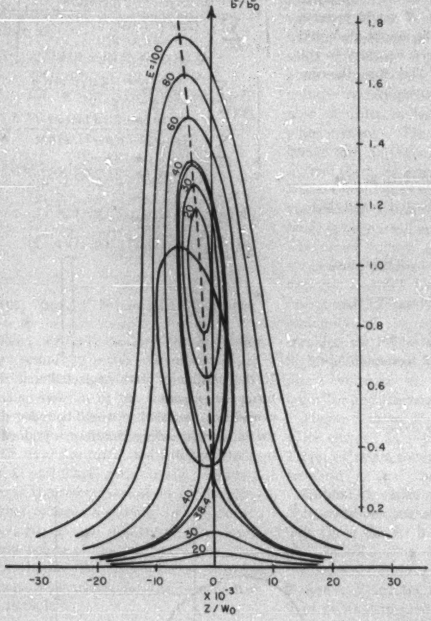


Figure 3 (left). Resonance curves and the threshold of stability, calculated from equations 19(A) and 22(A)

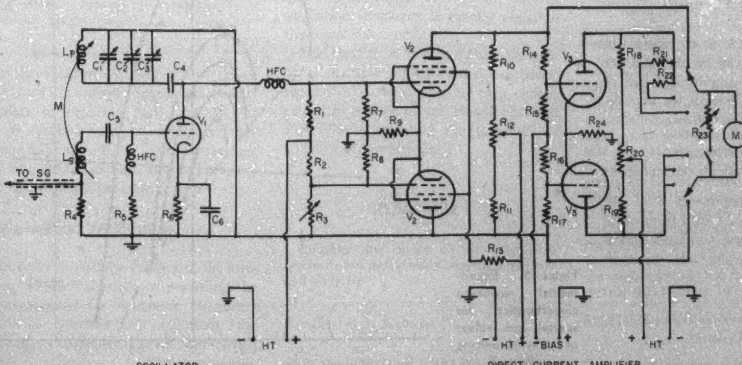


Figure 4. Connection diagram for the experimental determination of the resonance characteristics. Values of components are: C_1=1,000, C_2=10, C_3=5, C_4=C_5=500, C_6=100, R_1=R_7=R_{21}=47, R_2=100 (fixed)+50 (variable), R_4=R_9=1, R_5=12, R_6=5.6, R_7=R_8 = R_{22} = 470, R_{10} = R_{11} = 120, R_{12}=20, R_{13}=R_{14}=R_{15}=R_{16}=R_{17}=270, R_{18}=R_{19}=15, R_{20}=10, R_{23}=15 ohms, R_{24}=1.2; capacitances are in micromicrofarads and resistances in kilohms

$$b^2\left[\left\{s+\omega_0\sum_{r=2}(r^2-1)b_r^2/b^2\right\}^2+(\alpha_1-G)^2\times\left\{1-\frac{\displaystyle\sum_{p=3,5\ldots}[pa_p/2^{p-1}\{(p-1)/2\}!]\times\{(p-1)(p-2)\ldots\text{ up to }(p-3)/2\text{ factors}\}b^{p-1}}{\displaystyle\sum_{p=3,5\ldots}[pa_p/2^{p-1}\{(p-1)/2\}!]\times\{(p-1)(p-2)\ldots\text{ up to }(p-3)/2\text{ factors}\}b_0^{p-1}}\right\}^2\right]=\omega_0^2E^2 \tag{19}$$

It is clear from equation 19 that b is maximum when

$$s+\omega_0\sum_{r=2}(r^2-1)b_r^2/b^2=0 \tag{20}$$

indicating that the resonance curve is not symmetrical. The asymmetry is due to the harmonics present. If the harmonics were neglected and if terms up to a_3 were only considered, equation 19 would be reduced to that obtained by Van der Pol.[4] Again, at the stationary state, equation 16 becomes

$$v=b\left[\cos\omega t-(1/2\omega C)\sum_{r=2}\{G_1/(r^2-1)\}\times(h_{r-1}-h_{r+1})\sin r\omega t\right]=b\cos\omega t-\sum_{r=2}b_r\sin r\omega t \tag{16A}$$

where

b_r = amplitude of the rth harmonic $=(b/2\omega C)\{G_1/(r^2-1)\}(h_{r-1}-h_{r+1})$

G_1h_r is the value of G_th_r obtained from equation 5(A) after substituting b for b_t. Therefore

$$b_r=(1/2\omega C)\{1/(r^2-1)\}\times\left[(ra_r/2^{r-2})b^r+\sum_{p=(r+2),\,(r+4)\ldots}\times[rpa_p/2^{p-2}\{(p-r)/2\}!]\{(p-1)(p-2)\ldots\text{ up to }(p-2-r)/2\text{ factors}\}b^p\right] \tag{9B}$$

Putting $r=2$, 3, 4, 5, 6, et cetera, the different harmonics can be obtained, see appendix III. The amplitude b can be obtained from equation 19 for any detune z.

Stability of the Stationary Amplitudes of Forced Oscillations

The stationary amplitudes b of the fundamental can be obtained from equation 19. Whether these amplitudes occur in practice or not depends on their stability. Now

$$b_t=A_s\epsilon^{\int X\,dt}$$

where

$$X=(1/2C)G_t\{(h_2/2)-1\}$$

G_th_2 is to be obtained from equation 5(A) by putting $r=2$. Near the stationary state, $X\to 0$ and

$$b_t(=A_s\epsilon^{\int X\,dt})\to b$$

At the stationary state, $X=0$ and $b_t=b$. Therefore, the condition for stability could be stated as follows: if there is a slight increase δb in the possible stationary amplitude b (given by equation 19), then the increase δX in X must be negative, that is X must be reduced from its stationary value which is zero to a negative value, otherwise $A_s\epsilon^{\int X\,dt}$ would increase beyond b and hence the amplitude would rise indefinitely which is contrary to the fact, and vice versa. Therefore, the condition for stability is given by

$$(dX/db)<0$$

that is

$$(d/db)[(1/2C)\{G_t((h_2/2)-1)\}_{b_t=b}]<0$$

Using equations 3(A), 4(A), and 5(A), $(dX/db)<0$ becomes

$$\left[\omega_0C(E/b)\cos\phi\,d\phi/db-\omega_0C(E/b^2)\sin\phi-\sum_{p=3,5,7\ldots}(p-1)[pa_p/2^{p-1}\{(p-1)/2\}!]\{(p-1)(p-2)\ldots\text{ up to }(p-3)/2\text{ factors}\}b^{p-1}\right]<0$$

Differentiating 17(A) gives

$$\omega_0E\sin\phi\,d\phi/db=z+\omega_0(d/db)\times\left\{\sum_{r=2}(r^2-1)b_r^2/b^2\right\} \tag{21}$$

Substituting equations 17(A), 18, and 21 in the inequality gives the condition of stability at low synchronizing signals which is

$$\left[\left(z+\omega_0\sum_{r=2}(r^2-1)b_r^2/b^2\right)\left[z+\omega_0(d/db)\times\left\{\sum_{r=2}(r^2-1)b_r^2/b^2\right\}\right]+\{(\alpha_1-G)/C\}^2\times\left\{\frac{\displaystyle\sum_{p=3,5\ldots}[pa_p/2^{p-1}\{(p-1)/2\}!]\times\{(p-1)(p-2)\ldots\text{ up to }(p-3)/2\text{ factors}\}b^{p-1}}{\displaystyle\sum_{p=3,5\ldots}[pa_p/2^{p-1}\{(p-1)/2\}!]\times\{(p-1)(p-2)\ldots\text{ up to }(p-3)/2\text{ factors}\}b_0^{p-1}}-1\right\}\times\left[\frac{\displaystyle\sum_{p=3,5\ldots}[pa_p/2^{p-1}\{(p-1)/2\}!]\times\{(p-1)(p-2)\ldots\text{ up to }(p-3)/2\text{ factors}\}b^{p-1}}{\displaystyle\sum_{p=3,5\ldots}[pa_p/2^{p-1}\{(p-1)/2\}!]\times\{(p-1)(p-2)\ldots\text{ up to }(p-3)/2\text{ factors}\}b_0^{p-1}}+\{b/(\alpha'-G)\}\sum_{p=3,5\ldots}(p-1)\times[pa_p/2^{p-1}\{(p-1)/2\}!]\{(p-1)(p-2)\ldots\text{ up to }(p-3)/2\text{ factors}\}b^{p-2}]\right]\right]>0 \tag{22}$$

If the harmonics were neglected and if terms (in i_p-series) up to a_3 were only considered, equation 22 would be reduced to the corresponding condition obtained by Van der Pol.[4]

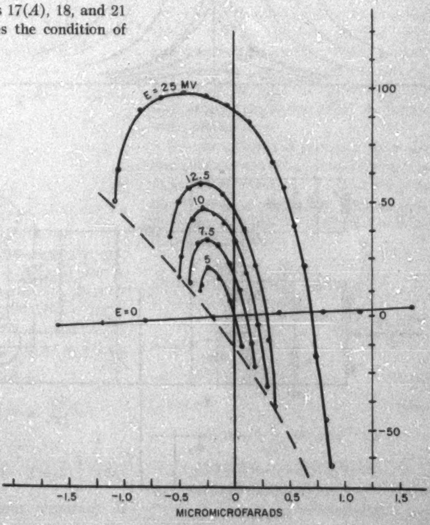


Figure 5. Experimental resonance characteristics for several low values of synchronizing voltages E

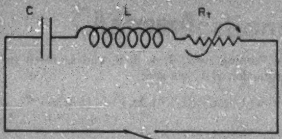


Figure 6. Series oscillatory circuit having a periodically varying resistance

If terms up to a_4 in the series of i_p are only considered, then equations 19 and 22 reduce to

$$b^2[[(z/\omega_0)+\{1/(\omega_0C)^2\}\{(1/3)a_2{}^2b^2+ \{(81/1024)a_3{}^2+(2/3)a_2a_4\}b^4+ (21/60)a_4{}^2b^6\}]^2+\epsilon^2\{1-b^2/b_0{}^2\}^2]=E^2 \tag{19A}$$

and

$$[[(z/\omega_0)+\{1/(\omega_0C)^2\}\{(1/3)a_2{}^2b^2+ \{(81/1024)a_3{}^2+(2/3)a_2a_4\}b^4+ (21/60)a_4{}^2b^6\}][(z/\omega_0)+\{1/(\omega_0C)^2\}\times \{a_2{}^2b^2+\{(45/128)a_3{}^2+(10/3)a_2a_4\}b^4+ (147/60)a_4{}^2b^6\}]+\epsilon^2(1-b^2/b_0{}^2)\times (1-3b^2/b_0{}^2)]>0 \tag{22A}$$

where

$$\epsilon=(\alpha_1-G)/\omega_0C$$

Taking $\epsilon=0.01$, $\omega_0/2\pi=500$ kilocycles per second, $C=500$ micromicrofarads, $b_0=10$ volts, $a_2=90\times10^{-7}$, $a_3\cong2\times10^{-7}$ and $a_4=0.3\times10^{-7}$, the resonance curves as well as the threshold of stability (obtained by equating equation 22(A) with zero) are calculated from equations 19(A) and 22(A) respectively. Figure 3 shows the resonance curves for synchronizing voltages E equal to 100, 80, 60, 40, 38.4, 30, and 20 millivolts. The maximum points of the resonance characteristics lie on the dashed curve given by equation 20 which, for the present particulars, is

$$(z/\omega_0)\times10^3+(1/225)\{2.7b^2+1.8285\times 10^{-2}b^4+0.326\times10^{-4}b^6+0.01823\times 10^{-6}b^8\}=0$$

The threshold of stability is also shown in Figure 3. Substituting the given particulars in equation 9(B), we get (at $b=10$ volts): second harmonic per cent = 2.55 per cent, third harmonic per cent = 0.12 per cent, fourth harmonic per cent = 0.0663 per cent.

Experimental Verification

In order to test the validity of the foregoing theoretical relations regarding the asymmetry of the resonance characteristics of synchronized oscillators, the following experimental work was carried out.

The oscillator used is of the ordinary tuned plate type, see Figure 4. The tuned circuit comprises a coil L_p, a main tuning capacitor C_1, a trimmer capacitor C_2, and a calibrated trimmer one C_3. The signal used to synchronize the oscillator, obtained from the "Marconi Echo" standard signal generator (S.G.), is injected in the grid circuit across the resistance R_4. As the variation of the output voltage of a synchronized oscillator in the whole synchronizing range is small, especially for the small input signals used in the experiments, a need is felt for a method to measure only these small variations. The change in the d-c plate current is a measure of the output voltage variation. A bridge system is used in order to balance the no-signal plate current. The arms forming the bridge are R_1, R_2, a variable resistance R_3, and the plate resistance ρ of the tube. Balance is obtained by adjusting R_3. The variation of ρ during the synchronizing range is very small and can be neglected. The direct voltage across R_1 and R_2 is applied to a balanced d-c amplifier having negative feedback by means of cathode resistances R_9 and R_{24} to increase the immunity of the amplifier from drift troubles, see Figure 4. A well regulated high tension source is used to supply the direct voltages to the plates of the oscillator and the amplifier.

Figure 5 shows the change in the d-c plate current as read on a central-zero 100-0-100 micro ammeter (M) of the d-c amplifier against the variation of the calibrated trimmer condenser C_3 for several small synchronizing voltages E. The curve for $E=0$ (free oscillation) is also shown. The free oscillation frequency is about 500 kilocycles per second. Before applying the signal, C_3 is always kept at its zero-position while the trimmer condenser C_2 is adjusted till the free oscillation frequency is exactly equal to the signal frequency.

It is clear from Figure 5 that the resonance characteristics are not symmetrical; the dashed curve connecting their ends is evidently a part of the closed curve of Figure 4 representing the threshold of stability at low signals.

Appendix I

Consider the series circuit, Figure 6, where R_t is a periodically varying resistance and given by

$$R_t=R_0\left(1+\sum_{r=1}^{\infty}h_r\cos r\omega t\right)$$

while the inductance L and the capacitance C are constants. The differential equation for the charge Q across the capacitor is given by

$$d^2Q/dt^2+(R_0/L)\left(1+\sum_{r=1}^{\infty}h_r\cos r\omega t\right)\times dQ/dt+\omega_0{}^2Q=0 \tag{23}$$

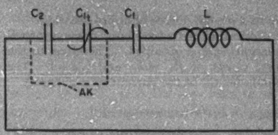


Figure 7. Connection diagram for equation 24

Substituting the transformation $Q=q\epsilon^F$ and choosing the function of time F such that dq/dt vanishes, and also neglecting terms containing high powers of $R_0/\omega L$ and $R_0h_r/\omega L$, equation 23 is reduced to

$$di_q/dt+\omega_0{}^2\left\{1+(R_0/2\omega L)(\omega/\omega_0)\sum_{r=1}^{\infty}\times rh_r\sin r\omega t\right\}\int i_q dt=0 \tag{24}$$

where

$$i_q=dq/dt$$

The function of time F is given by

$$F=-(R_0/2L)\left\{t+\sum_{r=1}^{\infty}(h_r/r\omega)\sin r\omega t\right\}$$

Let $\omega_0=\omega(1+d)$ and d = amount of detune, and it is assumed to be very small compared with unity.

The connection diagram giving the current iq of equation 24 will be as shown, Figure 7, where

$$1/C_1=\omega^2L,\ 1/C_2=(2d+d^2)/C_1,$$

$$1/C_{1t}=(1/C_1)(R_0/2\omega L)\sum_{r=1}^{\infty}rh_r\sin r\omega t$$

According to the well-known "compensation theorem," if short-circuiting C_2 and C_{1t} by closing the artificial key (a.k.) and replacing instead a generator of zero internal impedance and having an electromotive force equals $-\{(1/C_2)+(1/C_{1t})\}\int iq\,dt$, the actual current iq flowing in the circuit would be unchanged; but iq is unknown, therefore, successive approximations will be adopted as follows. Closing the a.k., the first approximation of iq would be given by $i_1=A\sin(\omega t+\phi)$, where A and ϕ are constants to be determined from initial conditions. For a second approximation, C_2 and C_{1t} are replaced by a generator of zero internal impedance and having an electromotive force $=v_1=-\{(1/C_2)+(1/C_{1t})\}\int i_1dt$. Closing the a.k. of Figure 7 and applying the component of v_1 of frequency $\omega/2\pi$, the current will be given by

$$i_2\cong A[\{(2d+d^2)/2\}\omega t\cos(\omega t+\phi)+ (R_0/\omega L)(h_2\omega/4)t\sin(\omega t-\phi)]$$

(i_1+i_2) gives the fundamental component of *iq* to a second degree of approximation.

Following the same lines, other approximations can be obtained. Therefore, to the first power of R_0, R_0h_r, and $(2d+d^2)$, the fundamental of the current *iq* is given by

$$i_{q\text{Fund.}}=B_1\epsilon^{\sqrt{a}t}\sin(\omega t+\phi')+ B_2\epsilon^{-\sqrt{a}t}\sin(\omega t+\phi'') \qquad (25)$$

where

$$\sqrt{a}=(\omega/)4\sqrt{(h_2R_0/\omega L)^2-\{2(2d+d^2)\}^2},$$

$$\tan\phi'=2(2d+d^2)/\{(4\sqrt{a}/\omega)+h_2R_0/\omega L\},$$

$$\tan\phi''=-2(2d+d^2)/\{(4\sqrt{a}/\omega)-h_2R_0/\omega L\}$$

B_1 and B_2 could be considered as constants depending on initial conditions. The term containing $\epsilon^{\sqrt{a}t}$ in equation 25 will be neglected as it is a highly damped one. The harmonics other than the fundamental are obtained as follows: replace C_2 and C_{1l} by a generator of zero internal impedance and having an electromotive force

$$\begin{aligned}v_h&=-\{(1/C_2)+(1/C_{1l})\}\int i_{q\text{Fund.}}dt\\ &=-B_1\{R_0/4\sqrt{1+a/\omega^2}\}\epsilon^{\sqrt{a}t}\times\\ &[h_1\cos(\phi'-\theta')+4(\omega L/R_0)(2d+d^2)\times\\ &\sin(\omega t+\phi'-\theta')+\sum_{r=2}^{\infty}-(r-2)h_{r-2}\times\\ &\cos\{(r-1)\omega t+\phi'-\theta'\}+rh_r\times\\ &\qquad\cos\{(r-1)\omega t-(\phi'-\theta')\}]\end{aligned}$$

where

$$\tan\theta'=\omega/\sqrt{a}$$

The effect of the component of v_h of frequency $\omega/2\pi$ is included in the predetermined fundamental, while the effects of the other components give the harmonic currents including the zero-frequency one.

Therefore, the current *iq* approximated to the first power of R_0, R_0h_r, and $(2d+d^2)$ will be given by

$$\begin{aligned}i_q=B_1\epsilon^{\sqrt{at}}\Big[&\sin(\omega t+\phi')-(R_0/4\omega L)h_1\times\\ &(\sqrt{a}/\omega)\cos(\phi'-\theta')+(R_0/4\omega L)\times\\ &\sum_{r=3}^{\infty}\{(r-1)/((r-1)^2-1)\}\times\\ &[(r-2)h_{r-2}\sin\{(r-1)\omega t+\phi'-\theta'\}-\\ &\qquad rh_r\sin\{(r-1)\omega t-(\phi'-\theta')\}]\Big]\end{aligned}$$

The charge

$$Q=q\epsilon^F=\epsilon^F\int iq\,dt$$

$$\begin{aligned}\cong(B_1/\omega)\epsilon^{(\sqrt{a}-R_0/2L)t}\Big[&\sin(\omega t+\phi'-\theta')-\\ &(1/2)(R_0/2\omega L)h_1\cos(\phi'-\theta')-\\ &(1/2)(R_0/2\omega L)\sum_{r=3}^{\infty}\{1/((r-1)^2-\\ &1)\}[(r-2)h_{r-2}\cos\{(r-1)\omega t+\phi'-\\ &\theta'\}-rh_r\cos\{(r-1)\omega t-(\phi'-\theta')\}]-\\ &(1/2)(R_0/2\omega L)\sum_{r=1}^{\infty}(h_r/r)\times\\ &[\cos\{(r-1)\omega t-(\phi'-\theta')\}-\\ &\qquad\cos\{(r+1)\omega t+\phi'-\theta'\}]\Big]\end{aligned} \qquad (26)$$

If R_0, h_r, and the amount of detune d are not constants but slowly varying functions of time, equation 26 becomes

$$\begin{aligned}Q\cong(B_1/\omega)\epsilon^{\int(\sqrt{a}-R_0/2L)dt}\Big[&\sin(\textstyle\int\omega dt+\phi'-\\ &\theta')-(1/2)(R_0/2\omega L)h_1\cos(\phi'-\theta')-\\ &(1/2)(R_0/2\omega L)\sum_{r=3}^{\infty}\{1/((r-1)^2-\\ &1)\}[(r-2)h_{r-2}\cos\{\textstyle\int(r-1)\omega dt+\\ &\phi'-\theta'\}-rh_r\cos\{\textstyle\int(r-1)\omega dt-\\ &(\phi'-\theta')\}]-(1/2)(R_0/2\omega L)\sum_{r=1}^{\infty}\\ &(h_r/r)[\cos\{\textstyle\int(r-1)\omega dt-(\phi'-\theta')\}-\\ &\qquad\cos\{\textstyle\int(r+1)\omega dt+\phi'-\theta'\}]\Big]\end{aligned} \qquad (26A)$$

It should be noted that the component of Q of frequency $\omega/2\pi$ is very approximately equal to

$$(B_1/\omega)\epsilon^{\int[\sqrt{a}-(R_0/2L)]dt}\sin(\textstyle\int\omega dt+\phi'-\theta')$$

Equation 23 is similar to equation 2 with the substitution of Q, R_0, L, and C for v, G_t, C, and L respectively. However, the fundamental of the output voltage v must be of the same form as assumed and which is $v=V_t\cos\omega t$, therefore, a possible solution for equation 2 can be obtained from equation 26(A) provided $\phi'-\theta'=\pi/2$, giving $d\cong-(1/8)(G_t/\omega C)^2h_2^2\cong 0$ (high powers of G_th_r being neglected), and $\omega\cong\omega_0$, therefore

$$\begin{aligned}v=A_s\epsilon^{(1/2C)\int G_t\{(h_2/2)-1\}dt}&\times\\ \Big[\cos\omega t-(1/2\omega C)\sum_{r=2}&\{G_t/(r^2-1)\}\times\\ &(h_{r-1}-h_{r+1})\sin r\omega t\Big]\end{aligned}$$

as given in the text.

No Discussion

Appendix II

Putting $r=2, 3, 4, 5, 6$, and so forth in equation 9(A) we get

$$V_2=(1/6\omega C)\{2a_2V^2+2a_4V^4+(15/8)a_6V^6+(7/4)a_8V^8+\ldots\}$$

$$V_3=(1/16\omega C)\{(3/2)a_3V^3+(15/8)a_5V^5+(63/32)a_7V^7+\ldots\}$$

$$V_4=(1/30\omega C)\{a_4V^4+(3/2)a_6V^6+(7/4)a_8V^8+\ldots\}$$

$$V_5=(1/48\omega C)\{(5/8)a_5V^5+(35/32)a_7V^7+(45/32)a_9V^9+\ldots\}$$

$$V_6=(1/70\omega C)\{(3/8)a_6V^6+(3/4)a_8V^8+\ldots\}$$

Appendix III

Putting $r=2, 3, 4, 5, 6$, and so forth in equation 9(B) we get

$$b_2=(1/6\omega C)\{2a_2b^2+2a_4b^4+(15/8)a_6b^6+(7/4)a_8b^8+\ldots\}$$

$$b_3=(1/16\omega C)\{(3/2)a_3b^3+(15/8)a_5b^5+(63/32)a_7b^7+\ldots\}$$

$$b_4=(1/30\omega C)\{a_4b^4+(3/2)a_6b^6+(7/4)a_8b^8+\ldots\}$$

$$b_5=(1/48\omega C)\{(5/8)a_5b^5+(35/32)a_7b^7+(45/32)a_9b^9+\ldots\}$$

$$b^6=(1/70\omega C)\{(3/8)a_6b^6+(3/4)a_8b^8+\ldots\}$$

References

1. THE INTERDEPENDENCE OF FREQUENCY VARIATION AND HARMONIC CONTENT AND THE PROBLEM OF THE CONSTANT FREQUENCY OSCILLATOR, **J. Groszkowski.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 21, 1933, page 958.

2. VALVE OSCILLATOR, **J. R. Tillman.** *Wireless Engineer* (London, England), volume XXIV, 1947, page 357.

3. THE APPLICATION OF POWER SERIES TO THE SOLUTION OF NON-LINEAR CIRCUIT PROBLEMS, **A. W. Gillies.** *Proceedings*, Institution of Electrical Engineers (London, England), volume 96, 1949, page 453.

4. NON-LINEAR THEORY OF ELECTRIC OSCILLATIONS, **Van der Pol.** *Proceedings*, Institute of Radio Engineers (New York, N. Y.), volume 22, 1934, page 1051.

5. A NEW METHOD FOR TREATING ELECTRON TUBES WHEN USED AS OSCILLATORS IN THE FREE AS WELL AS IN THE SYNCHRONIZED STATES, **Abd El-Samie Mostafa, M. El-Shishini Bey.** *AIEE Transactions*, volume 70, 1951, pages 999-1004.

Communications for Civil Defense

C. A. ARMSTRONG
MEMBER AIEE

HERE IN America, we have never been subjected to the kind of devastating all-out home-front attack that was launched against so many cities and towns of Europe, Asia, and Africa during World War II. There are those who may ask, what is civil defense and why does it need communications? Let us consider first, what civil defense is. Briefly, civil defense may be described as the protection of the home front by civilians, acting under civil authority, to minimize casualties and war damage and to preserve the maximum civilian support of the war effort. When a country is under threatened or actual attack, civil defense becomes a full partner of the military in the defense of that country. In recent years civil defense has grown in importance because the progress of air warfare has increased the possibilities of attacks on the civil populace. Industries, vital facilities, and population centers are all likely targets for hostile attack.

Now, why does civil defense need communications? There are two principal reasons. First, because civil defense is responsible for giving the people advance warning of impending attacks. This responsibility is shared jointly with the air-defense command of the United States Air Force in a manner which will be described in the following. The high speeds of modern aircraft require that the communications for alerting be speedy and efficient if the public is to receive adequate advance warning. This is a most important phase of civil-defense communications because, based on experience in Europe and Japan, it has been estimated that the casualties are reduced 50 per cent or more, simply by the giving of adequate advance warning of an air attack.

Secondly, communications are needed because civil defense has responsibilities in connection with fire fighting, rescue, medical, debris clearance, emergency repair or restoration of vital facilities, monitoring for specific hazards or special weapons, and so forth. For all these missions, communications are required to determine the situation at any moment, to dispatch forces, and to direct and coordinate activities toward the relief of that situation. Communications are needed between fixed points, also between fixed points and mobile units. Communications are needed within cities, between cities in the same state, and on a nation-wide basis. In a real sense, communications are the nerves of the civil-defense effort.

Engineering Requirements

The engineering of communications for civil defense must be based on the following factors

1. Coverage—able to interconnect all points between which communications are required and able to be expanded readily.
2. Expense—a minimum consistent with other requirements
3. Reliability—dependable under disaster conditions
4. Security—not susceptible to interception or unauthorized transmissions
5. Simplicity—easily operated by volunteers

Does civil defense need to build an extensive communications network for its own exclusive use in carrying out its mission? This question will be answered for many detailed instances in the following; however, at the outset, it is well to recognize a few salient factors. The communications requirements for civil defense might well be compared with the road and highway requirements for civil defense. Obviously, it would be impracticable to build a network of highways for use only by civil-defense forces and to be used only in time of disaster. Likewise, it would be impracticable to build a network of communications facilities for use exclusively by civil-defense forces in times of disaster. Like our road system, both our local and long-distance networks of communications facilities for normal peacetime usage are vast, diversified, and extensive. In the interest of providing the necessary coverage with adequate reliability and at reasonable costs, communications planning for civil defense needs to be based on the maximum utilization of existing facilities.

Attack Warning

Few people realize the extensive communications which lie behind the wailing of the air-raid warning siren. It is not simply the result of an observer somewhere seeing a flight of hostile aircraft and phoning someone to set off the sirens. If it were this simple, there probably would be many false alerts. But when the sirens sound an alert, it is the functioning of a carefully planned nation-wide system. This system involves several echelons whereby aircraft movements are reported, plotted, filtered, and analyzed by an Air Force general, or his deputy, who is stationed at an air-defense control center (ADCC). This general officer is responsible for the initiation of air-raid warnings throughout an entire air division, encompassing an area which includes several states.

AIRCRAFT DETECTION

Before attack warnings are disseminated, hostile aircraft must be detected and identified. To begin with there is the ground observer, who is located at observation posts spaced about 8 miles apart. In the ultimate plan, these posts will be uniformly spaced throughout the entire country. Using his ears and eyes, he scans the skies and reports to his filter center all the aircraft movements which he observes.

In general, regular telephone lines are used in transmitting the ground observers' reports to the filter center. However, these calls are accorded special handling. A distinctive phrase is used in placing these aircraft flash calls and this phrase immediately identifies these calls to the telephone operators, who handle them on a priority basis. Of course, aircraft flash calls are accepted only from officially authorized telephones. The traffic routing instructions are such that these calls are completed, not only to the correct filter center, but also to the correct position around the plotting table, where the call will be answered by a volunteer who is responsible for plotting reports from the particular observation post which is calling.

The ground observer reports the number of planes and type, their altitude, direction of travel, and other pertinent details. At the filter center the reports are plotted on a grid map and followed up as more reports come in. This information is put together and the course and position of the enemy aircraft are determined. In this way the enemy aircraft

---

Paper 53-187, recommended by the AIEE Wire Communications Systems and Radio Communications Systems Committees and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Southern District Meeting, Louisville, Ky., April 22–24, 1953. Manuscript submitted October 29, 1952; made available for printing March 2, 1953.

C. A. ARMSTRONG is with the American Telephone and Telegraph Company, New York, N. Y.

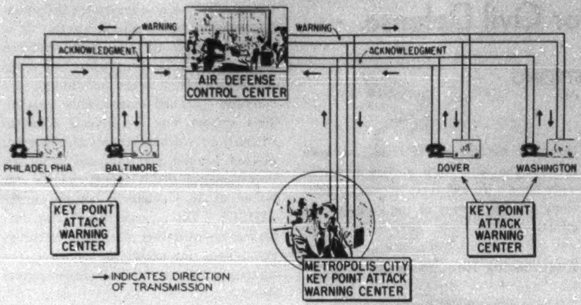

WARNING
ACKNOWLEDGMENT
AIR DEFENSE CONTROL CENTER
WARNING
ACKNOWLEDGMENT
PHILADELPHIA
BALTIMORE
DOVER
WASHINGTON
KEY POINT ATTACK WARNING CENTER
KEY POINT ATTACK WARNING CENTER
METROPOLIS CITY KEY POINT ATTACK WARNING CENTER
→ INDICATES DIRECTION OF TRANSMISSION

Fig. 1. Civil air-raid warning dissemination ADCC to key points using private-line voice network

is continuously kept track of and the probable targets are determined. Meanwhile, by means of private-line telephone circuits, the filter center relays this filtered track information to the appropriate radar plotting station. This track information is again plotted at the radar station, together with reports from adjacent filter centers. At the same time, strategically located radar sets are scanning the skies, locating and tracking any enemy aircraft within the range of their radar beams, and sending appropriate information, again by means of private-line telephone circuits, to the radar plotting station.

Thus, information about hostile planes flows into the radar station through two channels: 1. From the observation post, via the filter centers and 2. From the radar sets. The ground observer system is a vital part of the detection system and is needed to supplement the radar network because of the inherent limitations of radar. On the basis of all this information the radar plotting station will call the appropriate fighter-intercepter air craft bases into action; they send up their fighter planes to attack the enemy. These fighter planes are directed by the radar plotting station, which follows both our own and the enemy aircraft by radar and directs our fighter planes into the best positions for attack on the enemy aircraft.

By means of a private-line telephone circuit, the situation, with regard to hostile aircraft and any countermeasures taken by our aircraft, is continually being reported by all the radar stations within each division to the ADCC, which is the headquarters for the air division. At the ADCC, all of the information is continually plotted, and is analyzed by the commanding general, who is reponsible for the entire air division. This general is responsible for determining when and what type of alerts shall be disseminated throughout his division. When he determines that an alert should be transmitted, he passes this information to the civil air-raid controller at the center, who immediately disseminates the alert through the civil-defense attack-warning system.

Attack-Warning System

Alerts are disseminated from an ADCC to civil key-point warning stations by means of a special 4-wire private-line telephone circuit, which includes several unusual features. This civil air-defense warning (CADW) private-line circuit is illustrated by Fig. 1. A typical circuit would be 500 to 1,000 miles in total length and would connect an ADCC to about 25 key-point cities. Several such circuits may radiate from one ADCC. To send an alert, the controller at the ADCC operates a key which causes a bell to ring instantaneously and simultaneously in all of the key-point warning stations connected to this network. He then announces the alert, and this announcement is heard over loudspeakers, simultaneously at all the key points. The controller then calls the roll of key points for acknowledgments. The key-point attendant acknowledges using the push-to-talk hand set provided. The raising of the hand set (switchhook) cuts off the loudspeaker and permits conversation between the key point and the ADCC.

Several of the special features of this circuit are designed to prevent the origination of false alerts from any key point. Only one-way signaling is provided; only the ADCC can ring on this circuit. If a key-point operator wants to call the ADCC, he can speak into the hand set and his voice call will be heard over the loudspeaker at the ADCC, but no key

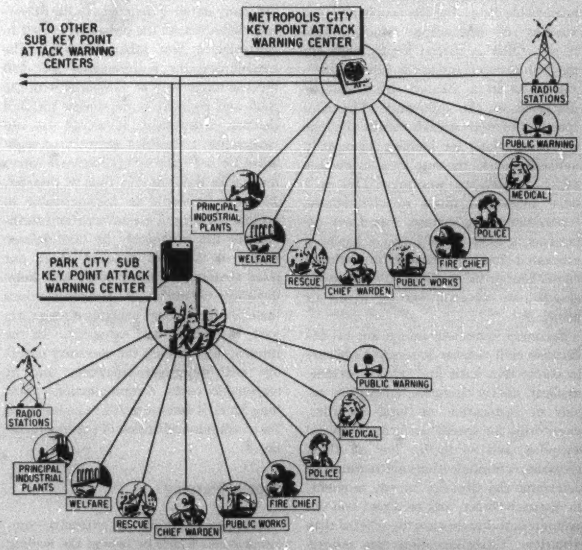

TO OTHER SUB KEY POINT ATTACK WARNING CENTERS
METROPOLIS CITY KEY POINT ATTACK WARNING CENTER
RADIO STATIONS
PUBLIC WARNING
MEDICAL
POLICE
FIRE CHIEF
PUBLIC WORKS
CHIEF WARDEN
RESCUE
WELFARE
PRINCIPAL INDUSTRIAL PLANTS
PARK CITY SUB KEY POINT ATTACK WARNING CENTER
RADIO STATIONS
PRINCIPAL INDUSTRIAL PLANTS
WELFARE
RESCUE
CHIEF WARDEN
PUBLIC WORKS
FIRE CHIEF
POLICE
MEDICAL
PUBLIC WARNING

Fig. 2. Civil air-raid warning dissemination key points to sub-key points, important warning stations, and the public by means of bell-and-lights network

Fig. 3. Operations room in a main control center. Representatives of each civil-defense service are around table and situation is plotted on wall map

Fig. 4. Message room in main control center. Booths with incoming telephone lines are in background and message evaluation table is in foreground

point can ring on the circuit. Furthermore, the two directions of transmission are completely separated so that the key points can not talk to each other; that is, the key points can hear only the ADCC while the ADCC can hear all the key points.

With this system, the alert reaches all the key-point personnel simultaneously so that the first step of CADW dissemination is completed in only about 10 seconds—the time it takes to ring the bells and announce: "Air-Defense Warning—Red, (Repeat) Air-Defense Warning—Red." The roll-call acknowledgement of the fact that the alert has been received and the degree of alert understood is completed in about 3 minutes, the exact time depending on the number of key points connected to a given network and the speed of response of the key-point personnel. A voice recorder is connected to this circuit at the ADCC so that a permanent record is available of this acceptance of responsibility by the key-point operator for the further dissemination of the alert.

Should the ADCC controller fail for any reason to receive an acknowledgment to the roll call from any key point, he will immediately place a long-distance telephone call to that point. For this purpose the ADCC has a specially provided toll-terminal telephone line. This is a direct telephone line to the nearest toll switchboard. A similar toll-terminal line is provided from each key point to the serving toll switchboard. In placing these calls, the controller uses a distinctive phrase which will, in an emergency, insure that the call receives the highest priority. With this arrangement, the entire network of toll message facilities (with many alternate routing possibilities) serves as a backup for the special CADW private line.

Key-point cities are chosen with consideration both to critical target areas and to geographical relationships. Some states have only one; others may have more than a dozen key points. Within the cities, the key-point warning station location is chosen as one which has reliable 24-hour coverage; for example, a police headquarters. From here the alert is disseminated to other points in the same city and to the adjacent cities and towns. For each target area it is desirable that an alternate location be designated for the key-point warning station. This alternate key-point location should be about 5 miles or more from the key point. It should be provided with a toll-terminal telephone line, which is served from a toll office other than the one serving the key point. The alternate key point receives alerts directly from the ADCC only if the regular key point is not operative, in which case the alternate will take over the functions of the key point.

The method used to disseminate alerts from key points varies, depending upon the following:

1. Existence of suitable private-line circuits which can be used
2. The number of points to be notified
3. The urgency, or speed requirements

This latter point may require some clarification, since one might argue that the dissemination of alerts is equally urgent to all points. However, as a practical matter speed is much more compelling in the alerting of points within or near target areas. There is an element of logic in the somewhat exaggerated idea that the weekly newspaper is a sufficiently good way to alert some of the remote rural areas in the interior of the country.

The need for an acknowledgment from the points being warned must be considered carefully. With many types of facilities, the cost is materially reduced if acknowledgments are not essential. In the first step of passing the warning (from the ADCC to key points) the cost of providing for acknowledgments is undoubtedly justified because of the importance of the key-point city and large numbers of people within each key-point area. In succeeding steps, as the number of people dependent upon the warned point becomes smaller, the number of points becomes very much larger and the possible over-all damage caused by a human or circuit failure becomes less. Since the cost of providing for acknowledgments increases very rapidly as the number of warning points expands, the acknowledgment feature usually cannot be justified in the lower steps of the warning chain.

To meet the requirements of alerting a large number of points quickly, simultaneously, and at a minimum cost, the Bell Telephone System has developed the "bell-and-lights" civil air-raid warning system. The use of this system is illustrated in Fig. 2. The system consists of a bell-and-lights indicator set installed at each location that is to receive the alerts; the set is so arranged that a different code ring and different light signal is received for each degree of alert. These indicator sets are all controlled by a special dial which is installed at the control point from which the alert is to be sent, normally a key point.

With the bell-and-lights system, a special sunflower-type dial is arranged with the different degrees of alert (rather than the usual letters and digits) shown

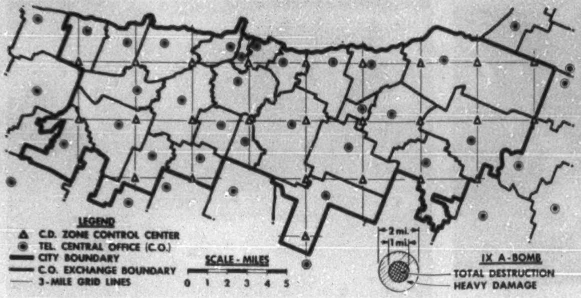


Fig. 5. Typical large city showing relationship of telephone central offices and zone control centers

by name and color on the petals outside the dial wheel. To send the yellow alert, for example, dialing the yellow color is all that is required. As soon as the dial action is completed, an audible code ring is received and a yellow light appears at those bell-and-lights stations which are authorized to receive the yellow (advance warning) alert. For the red (attack imminent) alert a red code ring and a red light appears at all bell-and-lights locations following the dialing of the red code. The white (all clear) is disseminated in the same manner by dialing the white code: a steady ring and a white light is transmitted. Provision has been made, in the design of this system, for a fourth alert—blue—which is not required at the present. By means of a simple strapping change in the central office equipment, the system may be so arranged that the yellow and/or blue can be sent either to all stations or only to a limited number of stations which have been previously designated. Any alerts sent from the master-dial location will continue until the control point operator dials "stop."

A larger number of warning points may be connected to a given bell-and-lights network. Alternate-control points may be provided, and generally should be at the same locations as the alternate key points. The principal circuits of the system are under continual automatic test so that trouble conditions will immediately register a major alarm in the telephone central office but will not cause the transmission of a false alert. This system employs dial-pulse signals between the control point and each central office. Between each warning station and its serving central office, the fundamentals of 4-party selective ringing are employed. The station indicator sets utilize a high-impedance ringer and cold-cathode neon lamps. One advantage of this system, as well as of the CADW private-line system, is that it is not dependent upon local power supply for operation; only telephone central office power is required. (As will be discussed later, these telephone offices are equipped with storage batteries and other emergency-power arrangements.)

The bell-and-lights system can be arranged to control the operation of public warning devices such as sirens, horns, whistles, and so forth. Public sirens which are connected to a bell-and-lights network will be sounded on the red and on the "public white" alerts simultaneously with the transmittal of the alert to bell-and-lights stations. The circuit is fully automatic and includes a memory feature which prevents the sounding of the sirens on confidential alerts—the yellow, the blue and the "confidential white" (a white which follows either a yellow or blue.)

In a number of cities and states the alerts are disseminated from key points to subkey points by means of an existing police-teletype network, which is normally used for other purposes. This is speedy and provides simultaneous alerting to a large number of locations. Existing police-radio systems are also used for this purpose. However, with both teletype and radio networks, the operation of the system is dependent upon the reliability of the local power at all the warning stations. Should the regular power fail for any reason, the system would not work unless emergency arrangements have been provided.

With both the ADCC-to-key-point special private-line circuit and the key-point-to-public bell-and-lights system, the transmission of alerts is practically instantaneous. With this combination, the only time delay between the decision at the ADCC to transmit a red alert, and the sounding of sirens in cities throughout the air division, some hundreds of miles away from the ADCC, is the time required for two simple operations: 1. The voice announcement of the alert at the ADCC and 2. The operation of the dial by the key-point operator. Assuming the controller and key-point operators are at their posts and act promptly, both these operations (not including the acknowledgment to ADCC) could easily be completed in less than 30 seconds. Allowing some time for slow answers and so forth, less than 2 minutes might be a good estimate of the time required.

If it were necessary to alert the entire country in less than 2 minutes, this could be done if all locations were provided with the private-line arrangements outlined in the foregoing. However, the costs could not be justified to provide facilities

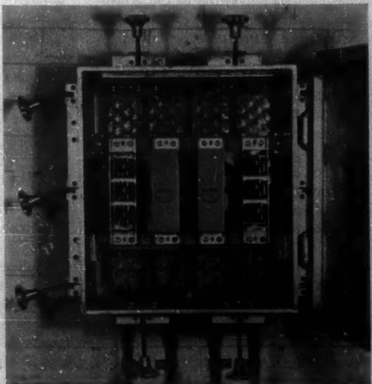


Fig. 6. Typical underground transfer relay installation. Steel door with moisture-proof seal has been opened and relay covers removed to show two banks of U-type relays

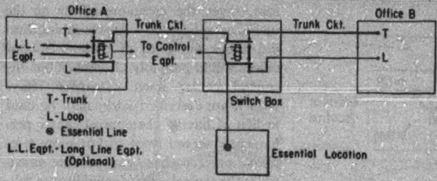


Fig. 7. Diversity protection obtained by means of transfer relay. Emergency operation of relay by remote-control transfers service from office B to office A

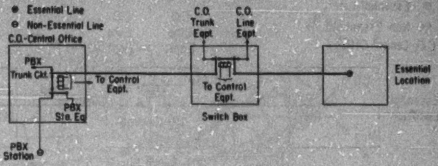


Fig. 8. Diversity protection obtained by use of PBX. Transfer relay in underground switch box may be remotely controlled to transfer important lines from normal telephone office to PBX terminations

so that the whole country, including every town and hamlet, can be alerted in 2 minutes. To alert those points, not in target areas, where the expense of a private-line system can not be justified, "chain" telephone calling is the most practicable method.

In planning to disseminate alerts by telephone calling, the chain must be carefully studied with due consideration to the number of telephone operators that may be on duty during periods of minimum traffic (such as midnight to morning) the number of telephone trunk circuits, and so forth. In some situations where it is important that the alert be received quickly at one or at a few points but where there are not enough such important points to justify the installation of a bell-and-lights system, it is practicable to install a special telephone which is used for no other purpose than receiving alerts. This is to avoid the possibility of an alert being delayed by a "busy" condition. The dissemination of alerts to the smaller cities and towns by means of chain calling can further be speeded if the calling is originated, not from a single point, but rather from a number of points all of which are served by a bell-and-lights network. In this way, the calling chain is built upon the fast fan-out of the bell-and-lights system.

Thus, the responsibilities for air-raid warning are shared. Civil-defense volunteers operate the observation posts and filter centers. However, the tactical control of the entire air-detection system, including observation posts and filter centers, and the manning of the radar stations and ADCC is the responsibility of the air-defense command of the United States Air Force. From this point on, civil defense is responsible for the dissemination of the alerts to key points, and from these to other cities, and so forth, and for the sounding of public sirens. The present aircraft-detection and attack-warning system which has been described is an outgrowth of World War II arrangements, and the experiences of Britain as well as of Germany and Japan. It is a tried and proven system. The principal change from World War II arrangements has been to increase the speed capabilities of the system, principally by the use of private-line circuits. This change has been necessitated by the higher speeds of the newer aircraft.

CONELRAD

Although not in itself an alert warning, Conelrad deserves a bried mention here because of its association with, and relation to, alerts. Very briefly, Conelrad, as its name implies—CONtrol of ELectromagnetic RADiations—provides a means of operating standard amplitude-modulation broadcast stations so as to deny usable navagational aid to an enemy. Other types of radio transmitters which might serve as navagational aids will either go off the air or will utilize special operational procedures during the period of the "Conelrad alert."

Many of the country's stations will receive their Conelrad alert by monitoring other key stations. Because of the small number of broadcast stations which operate on a 24-hour 7-day-a-week basis, and because of propagation problems, several hundred key stations must receive their alert notification from the ADCC by wire lines, in a manner somewhat similar to the arrangements previously described. The Conelrad alert may, or may not coincide either with the civil defense yellow or red alerts.

Disaster Control

Distinguished from attack warning, communications are required following an attack, for disaster control. Communications are used first in forwarding damage and casuality reports, so that authorities can obtain an over-all appraisal of the damage situation. Secondly, these communications are required to request, dispatch, and direct forces who will combat the spread of fires, rescue trapped people, control traffic, begin the restoration of essential vital facilities, and so forth. These communications connect together the various echelons of the civil-defense organization.

Within a given city, centralized over-all direction is required to control and combat the disaster situation; otherwise, resources might be wasted on hopeless or useless ventures. This over-all direction is the function of the main control center. An alternate main control center, located a considerable distance from the main control center, serves in the event of the destruction of the main control center.

Subordinate to the main control center, depending upon the size of the city, there may be several zone control centers. A zone control center serves its particular zone, generally including about 100,000 to 200,000 people, and about 10 square miles of area.

Within each zone there are usually five to ten report centers which receive damage reports from block wardens and other field workers and forward these reports to the zone control center. Any one (or several) of these report centers may serve as a command post for a damage control team, that is, it may serve as the field headquarters for a portion of the fire department, police department, the public-works department, and other civil-defense personnel.

There are a number of other locations where civil-defense forces will operate and will require communications. These include mobilization points, rendezvous areas where outside help assembles (mutual-aid and mobile-support teams), hospitals, radio broadcasting stations, and so forth.

MAIN CONTROL CENTER

Main control centers are selected in secure locations, generally the basement of a steel frame or reinforced-concrete building, and away from the most likely targets. For similar reasons main control centers need secure and reliable communications with other civil-defense forces. A main control center contains two princi-

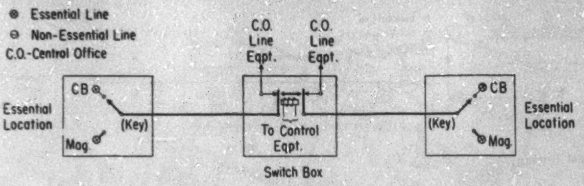


Fig. 9. Remotely controlled transfer relay in underground switch box used to connect essential locations

pal rooms which are important from a communications standpoint: an operations room and a message room.

In the operations room, which is illustrated in Fig. 3, a telephone is provided for the director of civil defense and for the head of each of the services such as police, fire, public works, transportation, and so forth. In addition to his regular local telephone, the director should also have a toll-terminal line. This will give him direct access to the long-distance switchboard. This will serve both to speed service for intercity and interstate communications and also to add an element of insurance of the continuity of communications.

A typical message room within a main control center is illustrated in Fig. 4. It is equipped with two groups of telephones. One group is used to receive incoming calls, and the other is used for the dispatching of outgoing messages. In the permanently setup control centers, where a refined arrangement is desired, the telephones may be of the operator-headset type. This frees both hands of the telephonist for writing and handling papers. With this arrangement, answering or placing a call (switchhook supervision) is accomplished by inserting the headset plug into a jack mounted on the table. To disconnect, the plug is removed. In dial-service areas, a permanently connected dial is mounted on the table. A separate bell or call indicator lamp (or both) is usually provided for each line.

The telephones in the message room are manned by volunteer telephonists who, in the case of incoming calls, record each message on a multicopy form. One copy is for each agency that may be concerned with any particular report. The telephones for incoming calls should be in a consecutive-number series, and appear in a "terminal hunting group"at the telephone office. This means that all of the phones in the incoming group will be assigned consecutive numbers. However, only one number, the number of the first line, need be used by those who call the control center. The first incoming call will then go to the first line, the next call will automatically switch to the second line if the first is busy, or to the third if both the first and second lines are busy, and so on to the end. The numbers assigned to the telephone lines serving a permanently setup control center should not be published and should be given only to those who may have a legitimate need to call the control center.

In addition to the foregoing, in the larger cities there might be provided in the message room a Teletypewriter Exchange (TWX) machine, for printed communication with state headquarters and the higher echelons of civil defense. This would be used primarily for forwarding summaries and situation reports and for requesting outside aid, material, and supplies. The larger cities might also make arrangements to have a teleprinter available for connection to a previously installed direct line to the Western Union switching center.

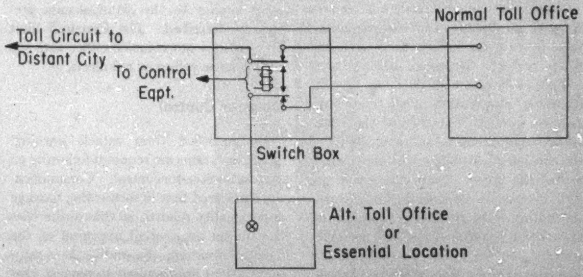
In order to insure the continuity of communications with the main control center, there should be telephone service from at least two central offices which are separated from each other by several miles. Under these conditions, a single bomb probably would not knock out both telephone offices and would not, therefore, sever the communications of the main control center. Any bombing that did knock out both of these telephone offices would probably also knock out the control center itself. Transfer relays (which are described later) may be used in lieu of having the control center permanently served from two or more telephone offices; these can provide essentially the same degree of reliability as full-time dual-office service.

The control center will usually be connected to the telephone offices by means of completely underground cable routes. Underground cables are little affected by an air-burst atom-bombing. These cables are already in their own "bombproof shelters." Whereas high-explosive bombs may sever any underground cables which they hit, this type of damage is of a localized nature and so, if adequate alternate routes are planned beforehand, it will not cut all communication paths with the control center. In the surburban areas, away from likely targets, the preference for underground rather than aerial cable is of less importance.

The foregoing arrangements (underground cable with multiroute and multioffice service) provide good insurance of the continuity of communications with the main control centers. These may be supplemented by a pool of the existing mobile radio services. Since this pool serves many other points besides the main control center, it will be described later in a separate section.

Zone Control Center

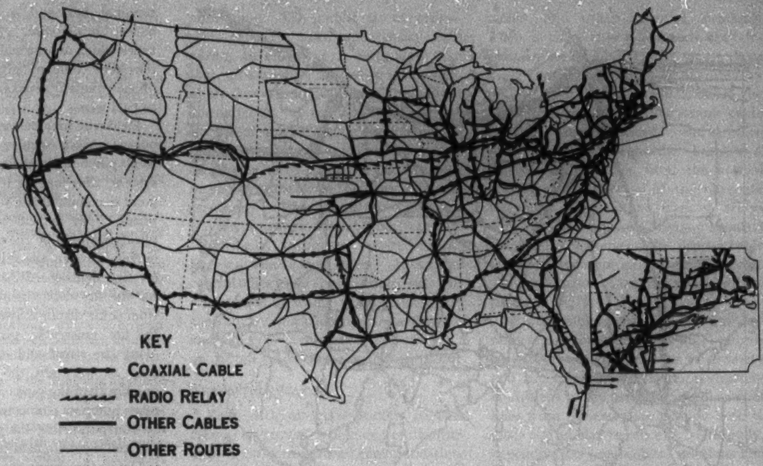
Whereas some cities have selected the zone control centers in municipally owned or rented space and equipped them on a full-time standby basis, this is a somewhat expensive plan. Therefore, many cities prefer simply to designate suitable loca-



Normal Toll Office
Toll Circuit to Distant City
To Control Eqpt.
Switch Box
Alt. Toll Office or Essential Location

Fig. 10. Diversity principle applied to long-distance service. Emergency operation of transfer relay in underground switch box will connect alternate toll office or essential location to toll switchboard in another city

Fig. 11. Principal toll routes of the Bell System and connecting companies



tions and facilities, which have normal peace-time functions and can serve in an emergency as a zone control center. The establishments selected to serve as zone control centers are those whose normal activities may be suspended or curtailed in the event of an emergency. The selection is based on the availability of existing communications and the desirability of obtaining a secure, well-protected location. A department store basement, or a protected location in a hotel are examples of such locations.

A compromise plan would include a few fully equipped zone control centers, which are used for no other purpose and which are employed for drills and training, together with a number of other zone control centers which have been designated in locations having other normal peace-time functions.

Similar to a main control center, but smaller, the zone control center also has an operations room and a message room. In the operations room, there should be a sufficient number of telephones, usually about 10, for the zone controller and for each of the civil-defense services represented. The message-room telephones are divided into incoming and outgoing groups just as in the main control center. Generally, about six phones are required for the incoming group and from two to four telephones for the outgoing-message group. While it is desirable to have the incoming phones connected in a terminal hunting group, this is not essential; any industrial or business establishment which has the required number of telephones may serve as a zone control center.

To understand the problem of reliability of communications for a zone control center, and to decide upon the extent of communications insurance needed, let us refer to a typical city. Fig. 5 illustrates the situation in a typical American city; the number of, and location of, the telephone central-office buildings as shown are identical with an actual city of the size indicated. (In New York City there are 60 separate telephone central-office buildings which are similarly dispersed.) Assume the worst type of attack such as an atom-bomb or a hydrogen-bomb attack; anything less would simplify the problem.

Referring to Fig. 5, suppose that an atom-bomb explodes, causing the destruction represented by the circle. In most of the locations that the circle can be placed, it will cover at least one zone control center and a telephone central office.

In certain places the circle may be placed to indicate the destruction of a zone control center but not its associated Central office, and vice versa. But there will always be an adjacent zone control center which is served from a working telephone central office; this adjacent zone control center can function as an alternate for a zone control center which either has been destroyed, or has lost its telephone service.

Therefore, it is generally not necessary to resort to any special measures to insure communications to a zone control center, as is done with the main control center. There may be a few exceptions which justify the use of transfer relays as discussed in a later section. Such an exception might be a city where there are a few, but an insufficient number of locations suitable to serve as zone control centers.

REPORT CENTERS

The report center is a very small communications center which serves as an intermediate level between field operations and the zone control center. From 30 to 50 warden posts may report to a single report center. In selecting existing establishments to serve as report centers, consideration is given to easy access, safety of construction, and availability of existing communications. Ideally, there would be three or four telephones arranged in an incoming-service group, plus one or two lines for outgoing service in forwarding evaluated reports to the zone control center. While it would be desirable to have the incoming lines in sequence, any establishment having a sufficient number of telephones may be utilized.

Report centers are generally located on a grid basis, roughly one report center at each intersection of 1-mile grid lines. Within each grid, three establishments should be selected which meet as nearly as possible the telephone requirements of a report center, with first, second, and third choices being designated. As a further aid to securing an organized group, the personnel of the establishments selected are often enrolled to serve as report-center personnel.

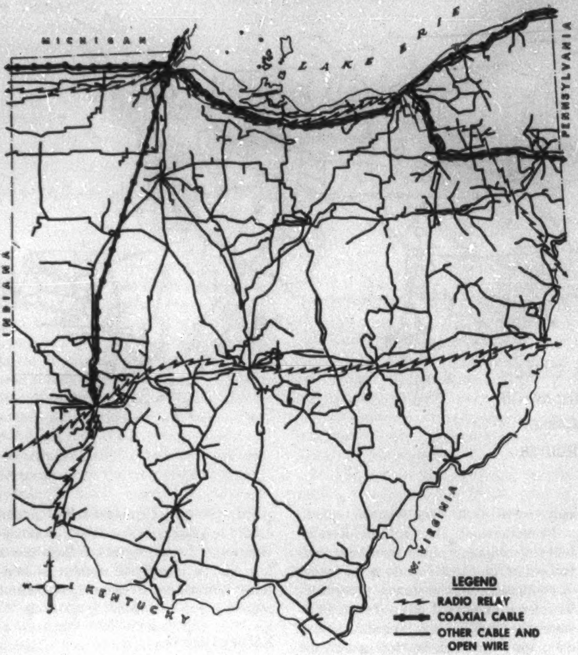


Fig. 12. Telephone toll routes in the state of Ohio

Miscellaneous Communication Points

In addition to the control and report centers, which are uniquely civil defense, there are many miscellaneous fixed locations to which civil defense will require a small number of communications channels. Such locations are the police and fire departments, most of the city departments, hospitals, each of the public utilities, transportation companies, radio broadcasting stations, assembly points, rendezvous areas, triangulation observation posts, and so forth. Generally, the existing telephone facilities to these places will provide adequate communications.

Where there are interruptions to the regular telephone service to any point, use may be made of mobile radiotelephones. This is a regular service provided by the telephone companies in 150 service areas, including nearly all of the cities classed as critical target areas by the Federal Civil Defense Administration. This service is provided on frequencies in the 152–162 megacycle range and in the 30–44 megacycle range. From a mobile radiotelephone, one may talk to any telephone in the city, or for that matter, any of the 45-million telephones in the country.

In addition to using the telephone company's mobile radiotelephone service, communications to any of these miscellaneous points may be supplemented by means of walkie-talkies or quickly run "field wire" and telephones. The hand-carried 2-way radio units are particularly useful where the points requiring communications are not fixed. They are somewhat less suitable for prolonged use between fixed points because of their battery-life limitations.

Field wire can be run within a reasonably short time interval to provide telephone service to fixed points. It may be carried by a man on foot or it may be unreeled directly from a moving vehicle. The term field wire as used here includes many types of wire which are not—strictly speaking—field wire, but which are similar and may be used for the same purposes. For example, "drop wire" which is normally used between a cable terminal box and a subscriber's residence, "bridle wire," "twist," and other types which are normally carried both in telephone company supplies and on telephone company trucks; all may be used as field wire.

Field wire may be used either for direct point-to-point circuits or to connect a fixed point with the general telephone system, thus making it possible for that point to communicate with any other telephone. One typical point-to-point application would be the establishment of a circuit between the relay pumpers used for fire fighting. When it is necessary for the fire department to pump water from some distance away to the scene of the fire, emergency pipe is quickly laid along the street and relay pumpers are used to maintain the pressure. More quickly than the pipe can be laid and the relay pumpers connected, field wire may be run paralleling the pipe from end to end and magneto field telephones connected on to the circuit at each pumper to be used for co-ordinating this tandem operation.

Pool of Existing Mobile Radio

In addition to the telephone company mobile radiotelephones, there are a large number of other existing 2-way mobile radio systems in everyday use. The police department, the fire department, and the taxicab 2-way mobile radio system are outstanding examples. In addition, there are ambulances, transit companies, the press, various industrial, and other mobile radio users. Further, there are a large number of amateurs whose help has always been valuable in emergencies.

Obviously, the fire and police radio systems will be utilized for the fire and police functions of civil defense. Rather than assign the other existing mobile ratio systems individually to specific uses (such as to assign the taxicab radio specifically to the warden service), it is generally preferable to assign to a pool all of the existing mobile radio systems which might appropriately be used by civil defense. Then in a disaster, units from this pool will be assigned to those locations where supplemental communications are needed. In this way, maximum use is made of the available mobile equipment.

Because none of the 2-way mobile radio systems (except the telephone company's mobile radiotelephone system) connect with the general telephone system, some additional arrangements must be made in order to get messages from the control centers to the various mobile units, for example, to the taxicabs. This may be accomplished by any of several means, such as the following:

1. The control center may simply telephone messages to the regular taxicab base station, at which point they will be relayed to the mobile units. This is the simplest and least expensive method.

2. By means of a remote control channel, between the control center and the regular base station, the control center may talk directly to the mobile units.

3. An alternate base station on the taxicab channel may be installed at the control center. This will permit direct conversation between the control center and the mobile units. This involves a relatively expensive alternate base station installation; sometimes the control-center location is a poor site for an antenna, which may result in poor transmission; in addition, arrangements must be made for emergency power.

4. As a variation of the foregoing alternate base station setup, prior arrangements may be made for one of the existing mobile units to serve as an alternate base station. This unit, equipped with two crystals and a channel-changeover switch, if necessary, may then be driven to the control center and used there as an alternate base station. Reasonably good transmission can generally be obtained if advance arrangements have been made for connecting the mobile unit to a previously erected antenna which is suitably located. This method avoids the installation of a separate base station and will be useful in the event that either the normal telephone service with the regular base station is interrupted or the regular base station is knocked out.

TELEPHONE LINE LOAD CONTROL

Since reliance is to be placed on the telephone system for essential communications at the time of a major disaster, means are provided to prevent the general public from overloading the switching facilities (switchboards and dial-switching equipment) by an excessively large number of calls being originated at the same time. The public is, of course, asked not to make unnecessary calls during an emergency. But more than that, measures are needed to insure definitely that calls from essential lines will not encounter an "all paths busy" condition. Protection against such a circumstance is made by the provision of line load control arrangements which insure that outgoing service is available to the lines which are essentially under disaster conditions, whether or not it is available to the general subscribers. Incoming service to all lines is available at all times. Authorized central office supervisory people, watching the volume of traffic through an office, can quickly and continuously adjust the load so as to avoid excessive overloads on the telephone central office equipment.

TRANSFER RELAYS

Used with alternate routes, transfer relays can provide a high degree of insurance of the continuity of communications in the event of an enemy attack. Transfer relays may be installed in underground locations, manholes, or cable vaults, and operated from a distant control point. Fig. 6 shows a typical installation of transfer relays mounted in a watertight metal case. In the installation shown, there are 12 relays which are capable of transferring a total of 24 circuits by means of one or more remote control circuits. The remote control circuits are under continual automatic test. Following are a few examples of how transfer relays may be applied to insure the continuity of service to preselected essential locations.

Transfer relays can be installed in a cable vault or manhole, near or beneath the regular serving central office, and connected to the lines serving a civil-defense control center, the fire and police departments, and other essential locations. Then, in the event of damage to the telephone office which normally serves these lines, these relays would be operated by remote control, and would transfer the lines (possibly via existing trunk cable facilities) to an alternate central office which is located in another building a considerable distance from the normal serving central office. At the alternate telephone office, other transfer relays which are associated with the same control circuit would connect the proper cable facilities to appropriate terminations. This application of transfer relays is illustrated in Fig. 7.

It will be observed that the transfer-relay arrangement shown in Fig. 7 affords service protection by means of diversity and, in this respect, is approximately equivalent to the permanent provision of lines from the essential location both to office *A* and also to office *B*. Generally, depending on geographical relations and other factors, the transfer-relay arrangement is less costly than the full-time provision of service from two offices. Furthermore, it will be noted that the cable facilities between offices *A* and *B* are normally used by the telephone company as interoffice trunks. These trunk cable facilities would be used exclusively by the essential subscriber only, in an emergency, in the event of an interruption in service from the regular serving office.

In those cities which have only one telephone building, transfer relays may be employed to connect the lines of the essential subscribers to a suitably located private branch exchange (PBX). This PBX may be one which is normally used for other purposes but which can be utilized for emergency operations. This is illustrated in Fig. 8.

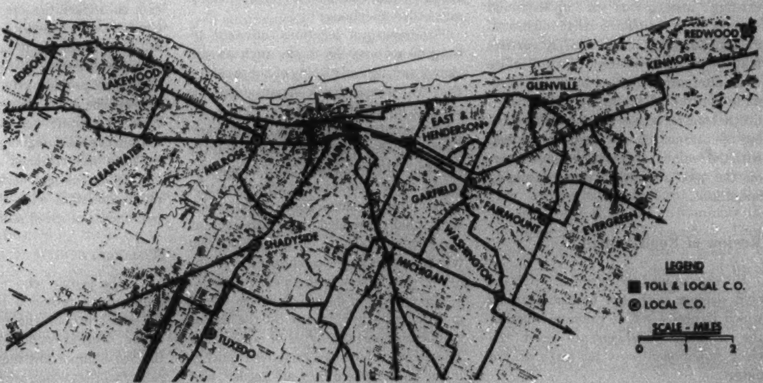


Fig. 13. Telephone central offices, toll and trunk facilities within city

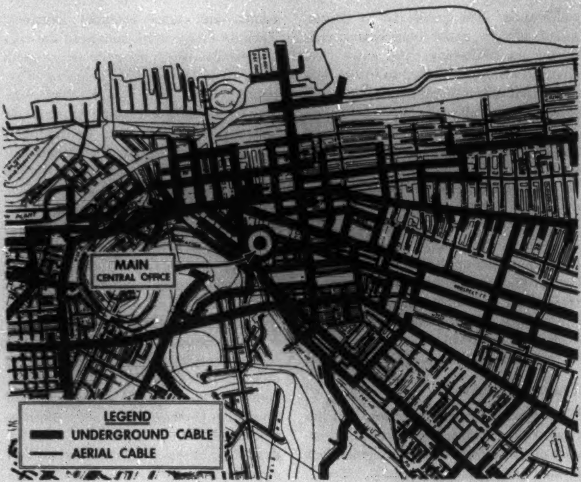


Fig. 14. Section of one central office area

If there is neither an alternate office nor a suitable PBX which can be used practicably in an emergency, then the transfer relays can be used to establish direct point to-point tie lines between the various essential locations. This is illustrated in Fig. 9. Transfer relays also may be used to facilitate the immediate restoration of service on various types of private-line circuits.

Transfer relays may also be employed to permit the immediate restoration of toll service to an alternate toll office or to emergency terminations in the event of the destruction of a toll office. This is illustrated in Fig. 10. With one of the existing installations of underground transfer relays similar to that illustrated in Fig. 10, 240 toll circuits (*K* carrier) may be transferred from one toll office to an alternate toll office, by means of remote control. Restoration plans which utilize transfer relays, in contrast to splicing operations, do not have to contend with the removal of debris, getting splicers on the job, and other time-consuming operations.

Review of Requirements

Having outlined the means of providing civil defense with the communications which it needs, let us examine whether the proposed arrangements meet the basic requirements, as set forth at the outset.

COVERAGE

The exact requirements for communications following an attack can not be completely foreseen. Therefore, maximum provision should be made for coverage, flexibility, and adaptability of civil-defense communications, both in respect to the points among which communications are available and also the volume of communications to be handled.

The regular telephone system inherently contains a great deal of coverage, in that some service exists in almost every building. There is complete flexibility in that any telephone or unit of the mobile telephone service can be connected to any other telephone.

For important locations, advance arrangements may be made, such as station prewiring and placing of house cables, that will permit quick expansion in the amount of telephone service. There are a large number of telephone-supply warehouses which are dispersed throughout the United States and which carry substantial stocks of various types of telephone material and equipment. These stocks, which are normally provided for growth—now at the phenomenal rate of about 2,000,000 telephones per year—can be used to establish or restore service in an emergency.

The maximum flexibility in the use of radio circuits is obtained by the use of mobile and portable types of equipment. These equipments can be moved quickly to the points where they are needed, and the specific circuits (channels) can be assigned to the most urgent use at the time. The proposed civil defense use of a pool of existing mobile-ratio units fulfills this need.

EXPENSE

In meeting their communications requirements, civil-defense officials find themselves in competition with other needs and agencies, both for money (federal, state, county, and city appropriations) and for materials, facilities, and equipment. In engineering a communications system for civil defense, one of the most important factors is the cost of alternative arrangements needed as insurance for continuity of service. In each case, studies are necessary to take into account all of the costs involved, since the initial outlay is not the only consideration. Costs of obsolescence of equipment must be considered, since it is entirely possible that civil-defense facilities will be needed for a long time. Maintenance costs usually are an extremely important factor if special facilities are to be provided.

Because the proposed arrangements would utilize, to the maximum extent practicable, existing facilities which have normal peacetime uses and would therefore, be paid for generally by the normal user, both the over-all cost and the cost to civil defense will be kept to the minimum. The cost of the transfer-relay arrangements as outlined in the foregoing, is generally less than the cost of other arrangements or facilities which would provide comparable diversity for protecting the continuity of service.

The principal situation in which special communications facilities must be provided by civil defense is for attack warning. Due to the very nature of the function these facilities are to perform, they cannot be "mothballed." Because they must be speedy and must be ready for instantaneous use, it is generally impracticable to provide attack-warning facilities on the basis that they will be disconnected until needed. The systems outlined in the foregoing for attack warning have been designed with the objective of keeping the long-term cost at a minimum consistent with the important requirements of speed and reliability.

RELIABILITY

To the people who will use the communications provided for civil defense, the reliability of these communications under disaster conditions is probably the most important requirement. Basically, facilities which are in normal day-today use are inherently more dependable under emergency conditions than are facilities which are provided for use only during an emergency. In keeping with this principle, the arrangements outlined in the foregoing are based on making the maximum use of all existing communications facilities, including both wire-line and radio facilities. Since communications are of paramount importance to civil defense, it is appropriate to examine carefully the factors which determine their reliability under disaster conditions.

An outstanding factor which contributes to the reliability of the regular telephone system for disaster communications is the great quantity and diversity of telephone plant. To appreciate this factor, first consider the nation-wide network; second, in effect place a magnifying glass over one portion of the nation-wide picture and see more closely the detail of the facilities within a given state; third, in effect place a higher-powered magnifying glass over an individual city within that state and study the number and locations of the individual central offices and the layout of interoffice trunks within that city; and fourth, in effect using a still more powerful magnifying glass, examine more closely the area served by an individual central office.

On a nation-wide scale, Fig. 11 shows only the principal toll routes of the telephone companies. The toll circuits connecting any two widely separated cities are normally routed, not all through the same cable and not even on the same route, but generally among several different routes. In addition to meeting the normal circuit requirements, this prevents any single incident, such as a bulldozer accidentally cutting a cable, from having any serious effect on the handling of normal day-to-day traffic. Furthermore, toll calls can be completed over alternate routes if their normal routing is cut. For example, a New York-San Francisco call which is normally completed via Chicago may be routed in an emergency via St. Louis, or via Atlanta and Dallas, or via any of several other routes. Rearrangement of the intercity circuits to meet normal daily fluctuations in demand for service, and to meet emergency conditions, is a regular day-to-day function of the telephone forces. On a nation-wide scale it is directed by traffic control bureaus located in Boston, New York, Washington, Richmond, Norfolk, Cleveland, Chicago, and San Francisco.

Not only is there diversity of routing available but also there is a diversity of the type of facilities used. There are ordinary toll cables, many employing type-*K* (12-channel) carrier systems; there are open wire lines, many employing type-*J* (12-channel) carrier systems; there are coaxial cables, all employing type-*L* (600-channel) carrier systems; and there are microwave radio-relay systems, each capable of carrying hundreds of telephone channels. Some of these cross sections are very heavy; for example, in the New York-Philadelphia cross section there are about 5,000 circuits which are divided among several different routes and different types of facilities. The Cleveland-Chicago cross section includes over 3,000 circuits, which are likewise diversified with respect to route and type of facility.

Having reviewed the nation-wide picture, let us place our magnifying glass over just one state. Fig. 12 shows the principal intercity toll routes within the state of Ohio; there are many routes shown here which could not be shown clearly on the nation-wide map. The same factors of diversity which exist on a national scale exist here on a state-wide scale. There are many other important factors, almost too numerous to mention, which have been provided to insure the continuity of communications. For example, toll cables are maintained under gas pressure (nitrogen, about 9 pounds per square inch), and are equipped with alarms at regular intervals. Any troubles which cause a break in the cable sheath will register a gas-pressure alarm in the nearest test room. By means of Wheatstone-bridge measurements on the alarm pair, the testmen can determine the approximate location of the trouble. Another example of emergency protective arrangements is the automatic line-switching circuit associated with the type-*L* carrier system; in the event of trouble on one line which causes as much as a 3-decibel change in the received level of a pilot channel, the line switching circuit will automatically transfer the service to a spare line. Similar arrangements are used on the radio-relay system to transfer service to a protective channel, in case of failure of the regular channel.

Now having seen the nation-wide and state-wide pictures, let us look at a single city within Ohio. Fig. 13 shows the locations of the telephone central offices and the interoffice trunk-cable routes within the boundaries of one city. There are 16 different central office buildings, each of these serving its own neighborhood and separated from each other so that if one of them happens to be near "ground zero" for an atom-bomb, the others are not likely to be hurt seriously. In this connection, most telephone buildings are built of heavy steel-framed or reinforced-concrete construction, to carry a floor load of 150 pounds per square foot, which means that they will generally remain standing as long or longer than most other buildings in the neighborhood.

Referring again to Fig. 13. it will be noted that four of the telephone buildings include toll-switching facilities. Thus, toll operations are dispersed. Additional toll-dispersion work is currently under way in many of the cities which are considered the most important toll-switching centers.

Let us look at one of these central office buildings. This is shown in Fig. 14, which completes our picture of the great quantity and diversity of the existing telephone system. It will be noted that cables are routed along nearly every street. The important point here is that they are nearly all underground. In a sample of cities large and small, which are rated as critical target areas, over 80 per cent of all the wires are underground. In the vital downtown areas of the larger cities, this percentage is close to 100 per cent.

Another important factor in the reliability of telephone service is that the local station equipment (the telephones themselves) do not require any local power. The power for their operation comes from the telephone central offices; these are equipped with storage batteries of sufficient capacity to carry the load for a reasonable period of time. In addition, there are standby diesel or gasoline-engine-driven generators which can charge the batteries in the event of an emergency requiring their use. For use in the smaller offices which do not justify a permanently installed standby generator, there are portable generators which can be used as required.

The telephone companies have maintained for a number of years a quantity of emergency radio facilities which can be used quickly to restore service around breaks in individual lines. These units are strategically dispersed throughout the country.

A further important factor behind the reliability of telephone service is the support organization, with standardized methods and materials, that is in being. There are nearly 700,000 trained people who are normally, in their day-to-day business, operating and maintaining the telephone facilities of the country. These people are organized into a number of individual companies, each one of which is a complete operating unit within itself. Often in past emergencies, earthquakes, floods, hurricanes, and so forth, it has been demonstrated how these trained people can be loaned from different places to a hard-hit area and how they can all work together as a team.

Security

It is important that the privacy of civil-defense communications be maintained. Possible panic-creating information must not become available to the general public. Also the enemy must be denied information that would be helpful to him. A direct private-line wire circuit provides the highest degree of privacy for communication between two points. The proposed arrangements for post-disaster communications afford the best security that could be provided without multiplying the cost manyfold.

For attack warning, security has a special significance with respect to the prevention of false alerts. It is for this reason that the special private-line arrangements which have been outlined in the foregoing need to be provided. This reason, plus speed, justify the use of the direct private-line civil air-raid warning arrangements, in lieu of relying exclusively on existing facilities for attack warning.

Simplicity

Because civil-defense communications must be used largely by volunteer workers, it is important that their operation ,be kept as simple as possible to avoid excessive training requirements. Most people are familiar with the use of regular telephone service in their locality. It is simple and straightforward.

With radio circuits, simplicity of operation indicates that use should be made of voice radio rather that radiotelegraph. Practically all of the existing radio systems which will be included in the "pool" are voice operated. In the civil-defense plans to operate existing radio facilities, it is preferable to utilize the personnel who operate them in normal day-to-day business. They are most familiar with the facilities and with their capabilities.

Conclusions

Civil defense can best be made effective if it is organized and operated on the principle that existing organizations and existing facilities are used wherever they exist. Wherever possible, the functions of civil defense are assigned to existing organizations whose normal peacetime operations include such functions, or are related to them. This insures the maximum efficiency and competence of both personnel and facilities.

In keeping with the foregoing principle, the police department, the fire department, the department of public works, the water supply and sanitation departments, the health department, and hospitals are all assigned the task of training themselves for their emergency role in civil defense. To them is given the task of organizing and training the auxiliaries they will need in times of disaster. Similarly, it is appropriate that public utilities, such as the power and communications companies, be brought into civil-defense planning and assigned responsibilities for enlarging or adjusting their normal functions to meet an emergency situation.

Civil defense must be accomplished by reliance on those who are doing, on an everyday basis (in local, state, and federal government, and in industry), work that is similar or comparable to that required in a civil defense emergency. The most efficient and reliable procedure is, therefore, to make maximum use of all existing communications agencies, together with their trained personnel. These are adequate, sufficient, and reliable.

No Discussion

Path Testing for Microwave-Radio Routes

R. D. CAMPBELL
MEMBER AIEE

IN PLANNING the construction of a microwave-radio relay system, it is of utmost importance that the sites for the terminal and intermediate stations be so selected that undesirable radio-propagation conditions are avoided. At the same time it is desirable, for reasons of economy to minimize the number of intermediate stations required, and to utilize the smallest towers which will give satisfactory radio transmission. In order to achieve these objectives great care must be exercised in the selection of the station sites.

In selecting sites, it is customary to determine the path profile, to establish the clearance of the center line of the path above the earth's surface. The path clearance can sometimes be determined with the aid of suitable maps and visual sighting methods, but contour maps in many cases have been found to be inaccurate; they do not indicate obstructions such as high trees and buildings, and atmospheric conditions are often not good enough for the use of visual sighting methods. Furthermore, such information needs to be supplemented with information about the nature of the surface of the earth along the path, particularly the possibilities for strong ground reflections which may adversely affect transmission. The possibility of reflections can be roughly assessed by an experienced observer after consideration of maps and inspection of the terrain. This procedure may be satisfactory in special situations where the terrain is obviously conducive to strong reflections, such as in the case on overwater paths, or in cases where the terrain is rough or covered with trees or sizable vegetation and is obviously not likely to cause strong reflections. However, in most cases it is not possible to determine by inspection whether ground reflections will prove serious, or what antenna heights should be selected for best results on a given path. For these reasons, where substantial sums are to be invested in a microwave system, such as that required for a backbone radio relay route, it is the usual practice in the Bell System to make special tests of microwave propagation on each path selected. It is the purpose of this paper to describe

---

Paper 53-191, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Southern District Meeting, Louisville, Ky., April 22–24, 1953. Manuscript submitted October 20, 1952; made available for printing March 2, 1953.

R. D. Campbell is with the American Telephone and Telegraph Company, New York, N. Y.

The author acknowledges the assistance given him in the preparation of this paper by A. W. Post and A. G. Oxehufwud of the Long Lines Department of the American Telephone and Telegraph Company, K. Bullington of the Bell Telephone Laboratories, and many others.

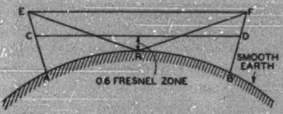


Fig. 1. Diagram showing direct and reflected rays—vertical scale greatly exaggerated

the equipment and procedures used in making such tests and to outline the method of analysis of the results obtained in typical cases.

Theoretical Considerations

The tests usually made to check path-propagation conditions consist in transmitting microwave energy between antennas at the ends of the path, elevated to various heights up to 200 feet above the earth. The purpose of these tests can best be understood by considering an idealized situation such as that indicated in Fig. 1, where *ARB* represents, in greatly exaggerated form, the surface of the earth along the desired path. For this case the earth is assumed to be perfectly smooth and to have a reflection coefficient approaching unity. With the antennas at *A* and *B* located at ground level, the direct path is obstructed by the earth and the radio energy is attenuated below the value calculated for free-space transmission over the same length path.

As the antennas are raised, the attenuation caused by the earth decreases until the signal strength reaches a value equal to that corresponding to transmission in free space. Theory and tests indicate that the free-space value is reached when the antenna elevations are such that the minimum clearance of the direct path above the earth is about 0.6 of the first Fresnel zone radius, as represented by line *CD* on Fig. 1. For antenna elevations below the line-of-sight path, only the small percentage of the energy which is diffracted beyond the horizon will reach the receiver. For all antenna elevations above the path barely grazing the earth's surface, there will be a reflected ray reaching the receiver in addition to the direct ray, and under the assumption of an earth with unity reflection coefficient, the reflected ray will have a strength equal to the direct ray.

With the antennas elevated above *CD* by an amount such that the length of the path traveled by the reflected ray *ERF* is between one-sixth and five-sixth wave length longer than the direct ray, the reflected wave will tend to reinforce the direct wave, and signal strengths above calculated free space values will be observed. The maximum received signal is 6 decibels above free space, and occurs at elevations such that the clearance of the direct path over the earth is equal to the radius of the first Fresnel zone. With this clearance, the path of the reflected ray is one-half wave length longer than the path of the direct ray, and the two waves add since there is a 180-degree phase shift on reflection. As the antennas are raised still higher the reflection path length will increase so that the reflected wave will arrive in such phase as to reduce the signal received over the direct path. When the reflected path is one wave length longer than the direct path, the two signals will just cancel each other. This condition occurs when the clearance of the direct path above the earth equals the radius of the second Fresnel zone. As the antennas are raised still further the received signal strength again increases, passing through another maximum 6 decibels above free space (clearance equal to radius of third Fresnel zone) and falls again to zero (clearance equal to radius of fourth Fresnel zone). This waxing and waning of the signal strength is repeated over and over as the antennas are raised, the maximum signals corresponding to clearances above earth equal to the radii of the odd numbered Fresnel zones, and the minimum signals corresponding to clearances equal to the radii of the even numbered Fresnel zones.

The dashed lines on Fig. 2 illustrate

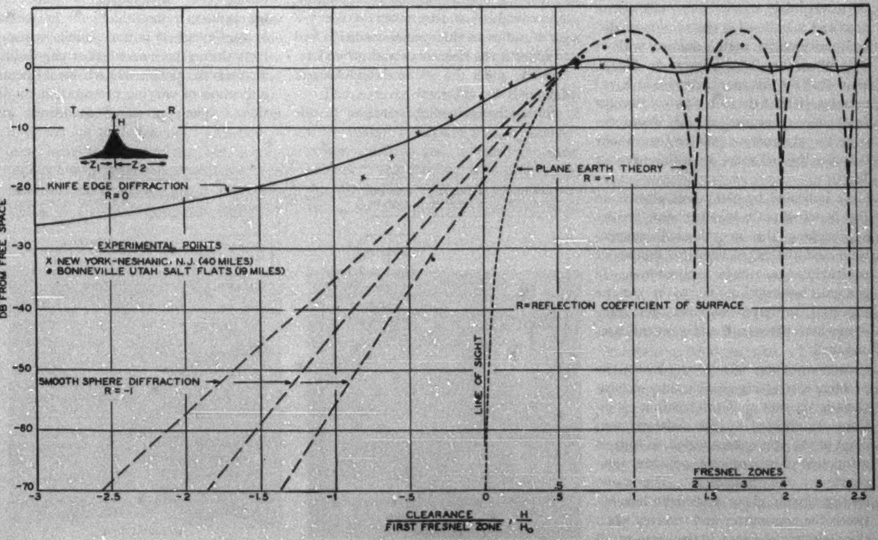


Fig. 2. Effect of clearance on radio transmission

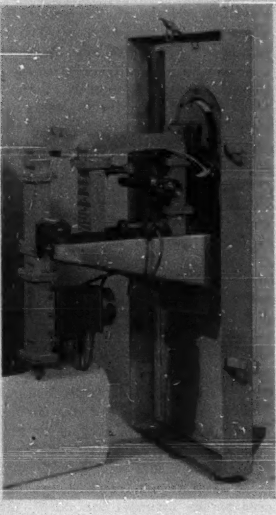


Fig. 3. Microwave transmitter for path testing

these theoretical considerations and show the calculated deviation from free-space transmission of signals received over smooth or plane earth, as a function of the path clearance. Calculations based on a path assuming a reflection coefficient of zero and a knife-edge obstruction or discontinuity on the earth's surface, such as that illustrated by the insert on Fig. 2, show that for this case the received signal remains within a decibel or two of that for free-space transmission, for clearances over the obstruction greater than about 0.3 of a Fresnel zone, see solid curve on Fig. 2.

As indicated by the points plotted on Fig. 2, actual paths having characteristics approaching the smooth-earth calculations and the calculations for knife-edge obstruction have been encountered. In general, however, actual paths can be expected to have characteristics lying somewhere between the theoretical cases assumed.

Another factor which complicates the problem of selecting suitable antenna locations is the fact that propagation conditions are affected by the dielectric constant of the atmosphere which, in turn, is a function of atmospheric pressure, temperature, and humidity. The radio energy travels along a straight line between the transmitter and receiver when the dielectric constant is the same at all elevations. When the dielectric constant increases with height, the radio waves are refracted upward and the path clearance of the energy that reaches the receiving antenna is less than the clearance for straight line transmission. Conversely, when the dielectric constant decreases with height, the radio waves are refracted downward, and the path clearance of the energy that reaches the receiving antenna is greater than that for straight-line transmission. Both conditions are encountered in practice.

For purposes of analysis it is often assumed that the radio energy travels in a straight line but that the earth radius changes to give increased or decreased clearance as conditions vary. This gives rise to the concept of "earth flattening" or "earth bulging." The degree of flattening or bulging is frequently expressed in terms of the ratio between the effective earth radius and the actual earth radius as, for instance, an earth of 4/3 or 2/3 actual radius.

It has not been possible to determine accurately what range of effective earth radii may be encountered in a given area, but estimates based on the rather meager information available indicate that the changes follow the laws of probability, varying around a mean value of about 4/3 radius with extreme conditions represented by "flat earth" (infinite radius) on the one hand, and 2/3 earth radius on the other hand, being reached only a few per cent of the time during the worst months. Accordingly it is customary to use 4/3 earth radius as the normal condition and to investigate clearances and ground reflections under the extreme conditions of flat earth and 2/3 earth.

Since the preceding changes in the transmitting medium have the effect of causing the effective clearance on any path to increase and decrease with time, its effect on transmission over the theoretical path assumed in Fig. 1 is to cause the characteristic curves of Fig. 2 to shift to the right or left along the abscissa as transmitting conditions change. Care must therefore be taken to pick antenna heights so as to avoid, in so far as possible, operation in the nulls caused by cancellation of the direct by the reflected signals, as the transmission characteristic shifts. In the majority of cases, reflections occur from several surfaces and the aim is to choose such antenna heights that the nulls from the various surfaces will not coincide as conditions change. Where strong ground reflections cannot be avoided, as sometimes happens on paths over water or over salt flats, it is considered good practice to select antenna elevations which fall, if possible on a broad hump of the characteristic between nulls. Elevations providing a minimum clearance of about 0.25 first Fresnel zone radius for the 2/3 earth radius condition to 1.7 for the flat-earth condition are considered about optimum for such paths.

It will be apparent from the foregoing discussion that path tests should, if possible, be made under normal atmospheric conditions, or transmitting conditions which are likely to prevail most of the time. Such conditions generally exist during the daylight hours when the atmosphere is most likely to be stable. In any event, it is undesirable to make tests during periods when strong fading is observed, as this in itself would be an indication of varying propagation conditions. The prevailing conditions are

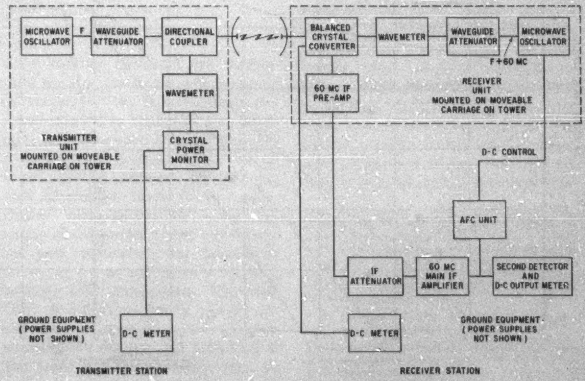

MICROWAVE OSCILLATOR
F
WAVEGUIDE ATTENUATOR
DIRECTIONAL COUPLER
WAVEMETER
CRYSTAL POWER MONITOR
TRANSMITTER UNIT MOUNTED ON MOVEABLE CARRIAGE ON TOWER
GROUND EQUIPMENT (POWER SUPPLIES NOT SHOWN)
D-C METER
TRANSMITTER STATION
BALANCED CRYSTAL CONVERTER
WAVEMETER
WAVEGUIDE ATTENUATOR
MICROWAVE OSCILLATOR
F+60 MC
60 MC IF PRE-AMP
RECEIVER UNIT MOUNTED ON MOVEABLE CARRIAGE ON TOWER
D-C CONTROL
AFC UNIT
IF ATTENUATOR
60 MC MAIN IF AMPLIFIER
SECOND DETECTOR AND D-C OUTPUT METER
D-C METER
GROUND EQUIPMENT (POWER SUPPLIES NOT SHOWN)
RECEIVER STATION

Fig. 4. Block diagram of equipment for path testing

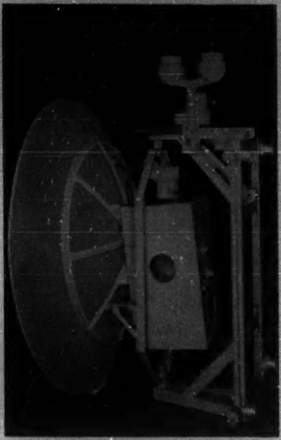


Fig. 5. Antenna carriage assembly

checked by observations on one or more obstacles in the path whose locations are definitely known.

Equipment Arrangements

The development and use by the Bell System of the Western Electric type *TD-2* radio equipment operating in the range 3,700–4,200 megacycles for broad-band telephone and intercity network television service in the United States has naturally led to the use of the frequencies in the same band for the path testing required.

Special microwave equipment has been designed for path testing. This equipment includes a transmitter, Fig. 3, containing a microwave reflex oscillator, type *406A*, which generates about 200 milliwatts of unmodulated radio-frequency energy. The equipment operates from a 60-cycle 120-volt a-c power supply. Arrangements are provided for adjusting the frequency of the transmitter to any frequency in the range 3,850 to 4,200 megacycles, but the frequency 3,950 megacycles is generally used, except where this might result in interference to existing systems. The output of the transmitter is fed to the antenna through a short wave guide system which includes an adjustable attenuator and a directional coupler. The directional coupler supplies a small percentage of the transmitter power to an absorption-type wavemeter for frequency measurement, and a crystal-detector unit for microwave output-power measurement. A block schematic drawing of the transmitter is shown in Fig. 4. The transmitter is arranged for mounting directly behind a 57-inch parabolic antenna, and the entire assembly is carried on a special framework, shown in Fig. 5. The antenna-transmitter assembly can be oriented either in azimuth or elevation by means of remotely controlled motors. The mounting frame is arranged with rollers designed to roll along the side legs of a portable aluminum tower, to facilitate raising and lowering the antenna-carriage assembly as required. The power-supply circuit to operate the transmitter, the connection from the output of the crystal monitor, a lead for adjusting the voltage on the klystron repeller to provide frequency control, and the circuits for controlling the orientation of the antenna assembly are extended in a cable to a power and control cabinet at ground level.

The receiver assembly of the microwave path-testing equipment consists of three units: a receiver unit, a receiver-power cabinet, and a receiver-control cabinet. The receiver unit is mounted directly behind a 57-inch parabolic antenna, the assembly being fastened to at carriage identical with the transmitter carriage in order to facilitate elevation to the desired height on the portable aluminum tower. The external appearance of the receiver unit is identical to the transmitter unit shown in Fig. 5. A view of the receiver with the cover removed is shown in Fig. 6. This unit includes a balanced crystal converter, in which the received microwave energy is combined with the output of a reflex oscillator, type *397A*, and associated variable attenuator, to produce a 60-megacycle difference frequency. This 60-megacycle frequency is amplified by a 3-stage preamplifier, the output of which is connected to the receiver-control cabinet (usually at ground level) through a 230-foot coaxial cable. At the receiver-control cabinet, the signal passes through a variable attenuator and the main amplifier, the output of which is rectified by a silicon detector, and connected to the signal-level meter which is calibrated in decibels above and below a zero reference. A high-impedance amplifier-limiter associated with a standard discriminator circuit is also bridged on the output of the amplifier to produce a control voltage, which is fed to the repeller of the klystron oscillator in the receiver unit to control its frequency so as to maintain the intermediate frequency constant at 60 megacycles. Fig. 7 shows the receiver-control cabinet. The receiver power cabinet which is shown in Fig. 8 contains regulated tube rectifiers for the three voltages required to operate the receiver assembly. The entire assembly is designed for operation from a 120-volt 60-cycle power supply. A block diagram of the receiving station is shown in Fig. 4.

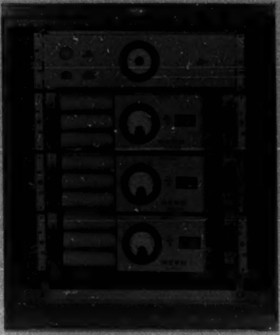


Fig. 8. Receiver power cabinet

As previously indicated, the transmitting and receiving antennas with their

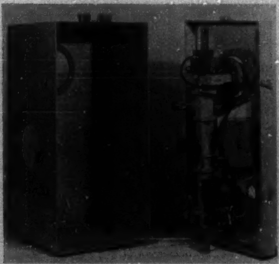


Fig. 6. Microwave receiver for path testing

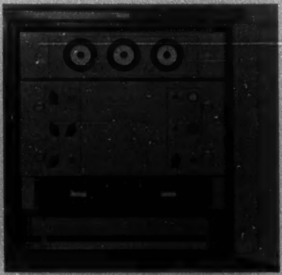


Fig. 7. Receiver control cabinet

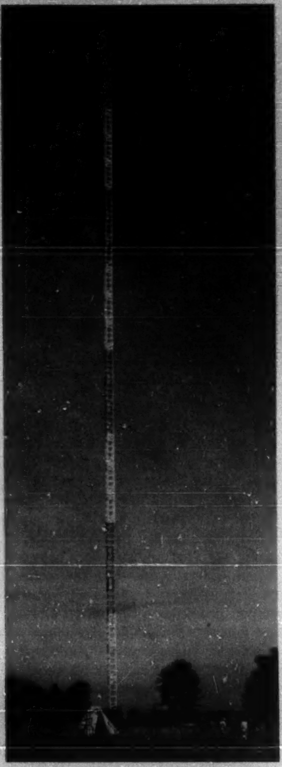


Fig. 9. 200-foot test tower

associated equipment are arranged for mounting on portable aluminum towers. These towers are fabricated of round aluminum-alloy members built up in sections 8 feet long and 22 inches square. These sections are fitted together with aligning pins, which make it possible to stack them one above the other, a maximum of 25 sections comprising a tower 200 feet high. A special erecting frame is provided which makes it possible to assemble the entire tower at ground level, successive sections being added at the bottom after raising the assembled portion by means of the erecting frame. Guys and winches are used to hold the tower vertical. Fig. 9 shows a 200-foot tower in place and Fig. 10 shows a receiver-antenna unit being hoisted up the tower.

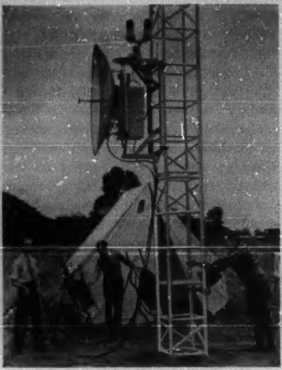


Fig. 10. Antenna-receiver assembly being hoisted up tower

Calibration of the equipment is carried out with the aid of measuring equipment, used first at the transmitting end and then at the receiving end of the test path. At the transmitter, the antenna-feed horn is disconnected and the energy at the output end of the transmitter wave guide is measured on a portable relative-power and frequency meter calibrated in decibels, referred to 1 milliwatt. The reading of the transmitter output meter is also recorded during this measurement, and the transmitter frequency is held constant throughout the subsequent path tests by means of this meter and adjustments of the repeller voltage. Any subsequent changes in output of the transmitter when it is properly tuned are noted and taken into account in the test data. The portable power and frequency meter is then used to check the receiver sensitivity, by measurement of the power supplied from a local signal generator adjusted to the test frequency and introduced through a calibrated attenuator into the antenna end of the receiver wave guide. The power is measured with the intermediate-frequency attenuator in the receiver adjusted to give reference reading on the signal-level meter. The path loss when receiving from the distant transmitter can then be determined from the calibration measurements and the intermediate-frequency attenuator setting at the receiver required for reference reading on the signal-level meter, proper allowance being made for the antenna gains, losses introduced by cable connections, and directional couplers in the test equipment.

The receiving and transmitting equipments for path testing each weigh approximately 2 tons. The majority of this weight is contributed by the tower, erecting frame, anchors, guys, and winches. The equipment for either station occupies about 1,000 cubic feet, and two medium-sized trucks and trailers are used to transport it. In addition, since the equipment must be operated at remote locations, portable emergency engine-driven generators must be used for power supply. Communication between the transmitting and receiving stations is usually obtained by the use of very-high-frequency portable radio sets.

Tests on Actual Paths

Two sections of the second New York-Philadelphia-Washington *TD-2* radio relay route have been selected for the purpose of describing in detail the results of path-testing measurements. Fig. 11 is a map showing the location of the two sections under consideration, namely, the section from New York (American Telephone and Telegraph Company Long Lines Building, 32 Avenue of the Americas) to a point near Red Bank, N. J., referred to as the Atlantic repeater station, and the section from Atlantic to New Egypt, N. J., about 13 miles east of Bordentown, N. J.

PATH FROM ATLANTIC TO NEW EGYPT

The Atlantic-New Egypt path, which is 23.6 miles long, was selected as typical of farming country with some hills: a type of terrain which is frequently encountered in the east in establishing intercity systems. A profile of the surface of the earth based on a 4/3 earth radius for this path is shown in Fig. 12. The ordinates show feet above mean sea level at New Egypt and Atlantic, and the abscissas, miles from New Egypt. Three series of test measurements were made over this path, transmitting from New Egypt on 4,130 megacycles, and receiving at Atlantic as described in Table I.

As indicated previously, test data were

Table I. Tests from Atlantic to New Egypt

Test	Description of Test
A.....	Measurements of path loss with the antennas at both ends elevated simultaneously in steps of 5 feet above ground, from 100 feet to 195 feet
B.....	Measurements of path loss with the New Egypt antenna fixed at 195 feet and the antenna at Atlantic elevated in steps of 5 feet above ground, from 75 to 195 feet
C.....	Measurements of path loss with the Atlantic antenna fixed at 195 feet and the antenna at New Egypt elevated in steps of 5 feet above ground, from 40 to 195 feet

obtained for maximum elevations using 200-foot towers. This procedure is followed in practically all such tests, except in the case of tests from building roofs or mountain tops where the area is too small to permit proper guying of the full tower. Data at the top of the towers frequently reveal obstructions near the center of the path which are not important when conditions are normal but which may become major obstructions under earth-bulging conditions. The results of the tests on the Atlantic-New Egypt path are shown on Fig. 13.

The limited height of the test towers (200 feet) did not permit obtaining data for clearances on this path greater than about the first Fresnel-zone radius. However, the fact that the maximum-signal strength observed was only about 1 decibel above free space indicates the absence of strong ground reflections on the path, under normal conditions, and subsequent studies did not reveal any likelihood of serious reflections under earth-bulging conditions. The general slope of curves *B* and *C* in the range just below free-space loss is of interest. The fact that the path loss decreases more rapidly with equal changes of antenna heights for curve *B* than for curve *C* is an indication that the controlling path obstruction for antenna elevations greater than 110 feet or so is nearer Atlantic than New Egypt (probably due to trees on hills *F* or *G*, Fig. 12).

From test *A*, Fig. 13 it is determined that the transmission loss is the same as the free-space loss, when the antennas at both ends are elevated 182 feet above ground. A straight line *A* has been drawn on Fig. 12 to show the center line of the transmission path with both ends at this elevation. As indicated in the discussion under the section on theoretical considerations, this indicates that the minimum-path clearance with the antennas at this elevation is approximately 0.6 Fresnel zone.

From test *B*, Fig. 13, it is seen that the crossover with free-space transmission occurs with the receiver at Atlantic elevated 182 feet and the transmitter at New Egypt elevated 195 feet. A line *B* interconnecting these points has also been drawn on Fig. 12.

Finally, from test *C*, Fig. 13 it is seen that the crossover with free-space transmission occurs with the transmitter at New Egypt at 145 feet, and the receiver at Atlantic at 195 feet. A straight line *C* interconnecting these points has also been plotted on Fig. 12.

By inspection of Fig. 12, it appears that the minimum clearance for path *C* occurs above point *D*. By plotting the 0.6 Fresnel-zone line corresponding to path *C*, it is found that its clearance over *D* is 33 feet and over *E* is 45 feet, as represented by the dotted arrows. Also a portion of the 0.6 Fresnel-zone line corresponding to path *A* has been plotted, and this line clears the earth at points *F* and *G* by about 45 feet (dotted arrows). Now if trees on hill *G* are the controlling obstruction, the transmission path would just graze the tree tops, with the New Egypt antenna at 195 feet and the Atlantic antenna at 118 feet above ground (see line *H* on Fig. 12). From curve *B* on Fig. 13, it is determined that the received field with the antenna 118 feet above ground at Atlantic is only 8 decibels below free space, and this appears unusually low. However, if trees on hill *F* are controlling, the grazing path would occur with the antenna at Atlantic 106 feet above ground (see line *J* on Fig. 12). From curve *B* on Fig. 13, the received field with the Atlantic antenna at 106 feet is found to be 10 decibels below free space, which appears reasonable. The controlling obstruction therefore appears to be trees about 45 feet tall on hill *F*. Similar analysis of the data with respect to hills *D* and *E* indicate that there are probably no trees or other obstacles on these hills.

In order to determine the antenna heights required, it is necessary to take into account the reduced clearances which may be experienced under extreme atmospheric conditions represented by an effective 2/3 earth radius. For this purpose a 2/3 smooth-earth contour has been plotted on Fig. 12. The separation between the contours measured along the ordinate at any point indicates the amount by which the clearance of the transmission path will be reduced when the conditions change from an effective earth radius of 4/3 to 2/3. Under the 2/3-earth condition, therefore, the tree tops at point *F* and the hill top at *D* would be pushed up from the arrow heads over these points to the points marked ×. If tower heights were selected so that the center line of the transmission path just cleared the highest of these points, the path would be just grazing under conditions of an effective earth radius of 2/3. Since the path loss at grazing clearance is seldom less than 10 or 12 decibels below free-space transmission, additional clearance is usually added to reduce this loss to about 4 decibels, for an effective earth radius of 2/3. In this example an additional clearance of 1/3 first Fresnel-zone radius has been added as indicated by the

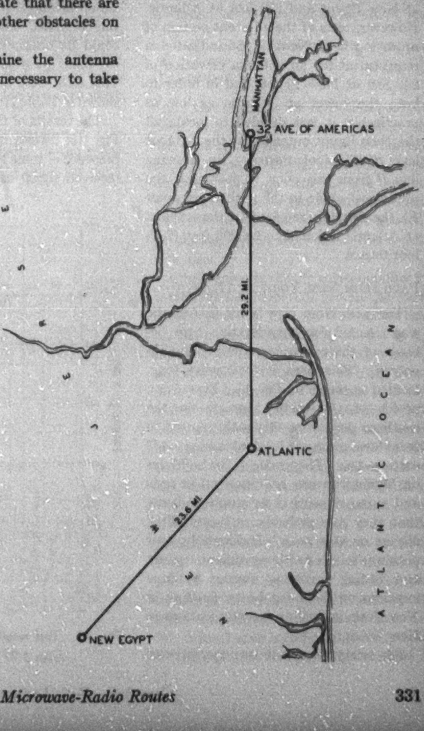


Fig. 11. Map showing test paths

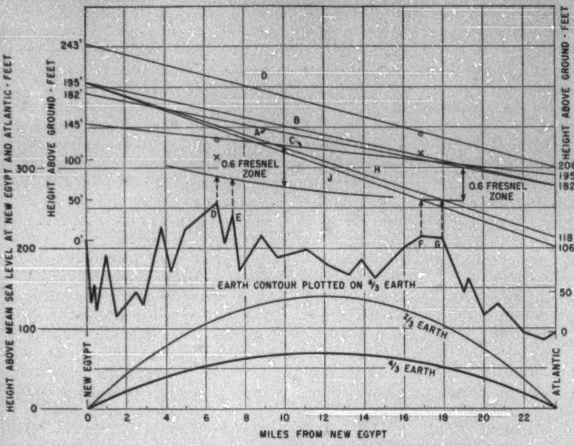


Fig. 12. Analysis of tests on New Egypt-Atlantic path, length 23.6 miles

Table II. Tests from New York to Atlantic

Test	Description of Test
B-8.....	Measurement of path loss with the transmitting antenna at New York fixed at 60 feet above the tower base and the receiver at Atlantic elevated in steps of 5 feet above ground
B-9.....	Measurement of path loss with the transmitting antenna at New York fixed at 15 feet above the tower base (on 28th floor of building) and the receiver at Atlantic elevated in steps of 5 feet above ground
C-3.....	Measurement of path loss with the receiving antenna at Atlantic fixed at 195 feet and the transmitting antenna at New York elevated in steps of 5 feet, from 10 to 60 feet above the roof
C-4.....	Measurement of path loss with the receiving antenna at Atlantic fixed at 120 feet and the transmitting antenna at New York elevated in steps of 5 feet from 10 to 60 feet above the roof

circles above the × points. A straight line passing through these circles would indicate tower heights of about 120 feet at New Egypt and 250 feet at Atlantic. However, tests of the adjacent section of route south of New Egypt indicated a requirement for an antenna elevation of 243 feet at New Egypt, and in order to keep the tower at Atlantic as low as practicable, it was decided to locate all the New Egypt antennas at the 243-foot level (237$^1/_2$-foot tower). By drawing line *D* from this point to just clear the uppermost circle in the path (over point *F*), the elevation required for the antenna at Atlantic is found to be 206 feet (200-foot tower).

Path from New York to Atlantic

The path from New York to Atlantic was selected for consideration here because of the overwater transmission involved. This path is 29.2 miles long, located as shown on Fig. 11. Fig. 14 is a profile of the earth's surface on the southern portion of this path, plotted in feet above mean sea level, assuming a 4/3 earth radius. The profile of the buildings on Manhattan was not prepared as tests and visual sighting observations indicate that they are probably not controlling factors on this path. Incidentally, this path was selected after considerable study and testing involving various antenna locations on the Long Lines Building in New York, as well as various locations in New Jersey.

For safety reasons it was necessary to limit the height of the test tower on the Long Lines Building to only 64 feet, and the test involving raising antennas simultaneously at both ends of the path was omitted, since no significant data could be obtained from such tests with the limited antenna height at New York. Test data were obtained on this path as shown in Table II.

The results of these tests are shown on Fig. 15. Tests *B-8* and *B-9* show the presence of some minor fluctuations of the received signal as the receiving antenna was elevated beyond about 80 feet, with a marked dip due to reflections at antenna elevations between 150 and 190 feet. The curves *B-8* and *B-9* on Fig. 15 cross the line representing free-space transmission at receiver elevations of 81 and 91 feet respectively. Lines representing the center line of the transmission paths at the free-space crossover points for tests *B-8* and *B-9* have been drawn on Fig. 14.

Lines *B-8* and *B-9* of Fig. 14 cross at about 4.8 miles north of Atlantic, indicating that the controlling obstruction is probably located on hill *D*. Six-tenths of a Fresnel zone for this location is 42 feet at the test frequency of 4,130 megacycles, and by plotting this on Fig. 14 it is determined that the obstruction extends about 30 feet above the top of hill *D*. Visual inspection indicates that trees located on this hill constituted the controlling ob-

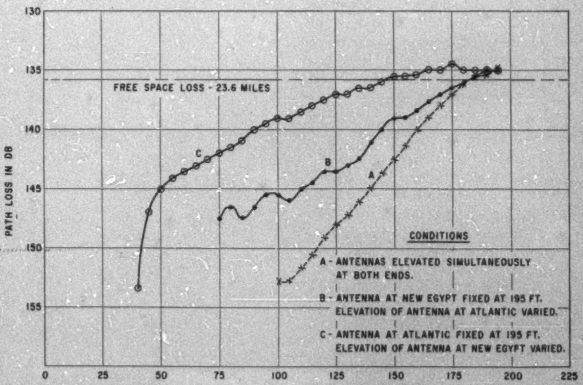


Fig. 13. Test results for Atlantic-New Egypt path, length 23.6 miles. Transmitter at New Egypt, receiver at Atlantic

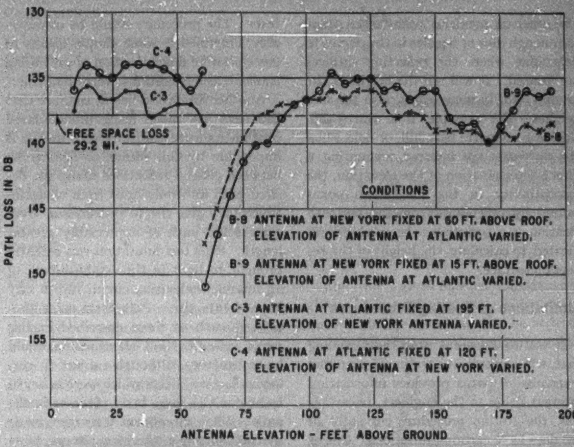


Fig. 15. Test results for New York-Atlantic path; length 29.2 miles. Transmitter at 32 Avenue of the Americas, New York City; receiver at Atlantic

struction in the path at the time of the tests. The hills marked *E* and *F* are also of interest, and the dotted arrows indicate the approximate height of trees which were found by inspection to be growing on these hills. Under earth-bulging conditions corresponding to an effective earth radius of 2/3, the tree tops on hills *D*, *E*, and *F* would reach the points designated × with respect to the ends of this path, and the circles are plotted to indicate elevations 1/3 Fresnel zone above the trees under this condition. Likewise the elevation of the top of the trees on these hills under flat-earth conditions are shown by the black dots below *D*, *E*, and *F* respectively.

As it was desired to avoid construction of a tower on the Long Lines Building, the elevation of the antenna on this end of this path was fixed at 459.5 feet which is 18.5 feet above the 28th-floor roof. Drawing a line from this point just above the circles over hills *D*, *E*, and *F* gives the minimum height required for the antenna at Atlantic as 145 feet above ground.

While the antenna pointing south from Atlantic was to be located 206 feet above ground, this elevation was obviously not suitable for the antenna pointing toward New York because serious reflections from the surface of Lower Bay would be experienced. In fact, reference to Fig. 15 indicates the presence of some reflections even at the 170-feet elevation. Hence it was decided to put the antennas pointing toward New York on a platform on the Atlantic tower at the 150-foot level, which makes the center line of the antenna 156 feet above ground. An antenna location at a somewhat lower elevation would be preferable, were it not for the obstructions on hills *D*, *E*, and *F* of Fig. 14. However, the tests plotted on Fig. 15 indicate that the transmission loss over the path will be about equal to free space under normal conditions (4/3 earth radius), using these antenna heights. With greater earth bulging, the water reflection path will be shielded even more, due to the trees on hills *D*, *E*, and *F*.

With a flat earth, the distance in miles from the Atlantic repeater to the reflection point is given by the following equation

$$d=\frac{h_1}{h_1+h_2}D$$

where

h_1 = height of Atlantic antenna above sea level

h_2 = height of New York antenna above sea level

D = length of path in miles

From this formula, the reflection point is found to be 10.3 miles from Atlantic. A line from the antenna location to this reflection point (*R* on Fig. 14) passes below the black dots indicating the tree tops on hills *E* and *F* under flat-earth conditions. Hence it appears that water reflection will be blocked by the trees under this condition also.

Consider again the dip corresponding to a receiver height of 170 feet in the path loss curve *B-9* of Fig. 15. The center line of this path has been plotted on Fig. 14 between the 170-foot elevation at Atlantic and the 456-foot elevation at New York. Now, by inspection it appears that the water reflection for this path under 4/3 earth radius conditions might take place at a point about 11 miles north of Atlantic. At this point, the center line of the direct path is measured on Fig. 15 as 235 feet above the 4/3 earth surface. The sixth Fresnel-zone radius for a 29.2-mile path is calculated as 230 feet, at 11 miles from one end of the path. This is considered to be a fairly good check on the measured clearance of 235 feet at 11 miles, and hence it is believed that the 6th Fresnel zone is responsible for the dip in the path loss curve. This method is

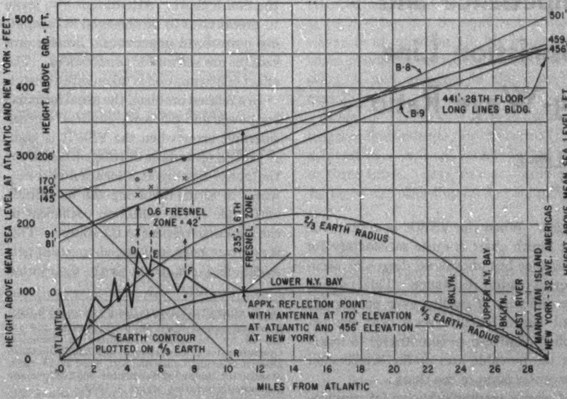


Fig. 14. Analysis of tests on Atlantic-New York path; length 29.2 miles

also used on overland paths when there are enough dips or wiggles in the curves to determine where the reflecting surfaces are located. It will tell the height of the ground and, in many cases, also the tree heights. For example, if a nearby hill shields a midpath, bare hill, or field when the antennas are lowered preventing a direct determination of the elevation, the discontinuity in the height-loss curve caused by reflections from the midpath location can sometimes be used as indicated to calculate the height of the reflection surface.

Limitations of Testing Procedures

It will be apparent from the foregoing that, while microwave testing of proposed transmission paths provides information of great value to the engineer, nevertheless, the testing procedure currently in use does have certain limitations. One of these is the fact that testing is normally limited for reasons of economy to relatively short periods of time, and hence the results do not show the magnitude or frequency of occurrence of variations in atmospheric conditions which may affect transmission. While the tests are usually made only at times when it is believed that transmission is normal, this is not always possible, and the test data must be carefully analyzed in all cases to be sure what conditions are prevailing during the tests. The procedure would be considerably simplified if some simple means of more positive determination of prevailing transmission conditions were available.

Another limitation lies in the fact that the tests are made using unmodulated continuous-wave transmissions, and it is impossible by this means to distinguish between ground reflections along the desired path and reflections from obstacles which may give rise to transmission over some other path of appreciably greater length. This test limitation was recently brought to light in the establishment of an intercity television circuit into a city of moderate size. Path tests using procedures such as those described in this paper indicated that transmission would be satisfactory, although subject to certain reflection effects which were believed to be due to a large body of water in the path. When subsequent tests were made by transmitting a television picture, however, a series of vertical black-and-white bars were observed which were found to be caused by energy overshooting the receiver to strike the face of a large building beyond, from which it was reflected to the face of another large building almost in the transmission path, and from there it was reflected directly into the receiving antenna. This double-reflection path had a length estimated to be some 12,000 feet greater than the direct path, and resulted in a signal delay of approximately 12 microseconds with respect to the main signal. Although the level of this reflected signal was considerably below that of the signal over the direct path, it was nevertheless sufficiently strong to seriously impair transmission, particularly during periods when fading was observed on the direct path. As a result of this experience, consideration is being given to the use of modified test equipment which will employ pulse or modulated-wave transmissions, with receiving equipment which is capable of measuring the amplitude and the time delay of reflected signals.

Conclusions

As the result of extensive observations and tests, it has been concluded that the use of testing techniques similar to those described here are of great assistance to the engineer in selecting sites and determining optimum tower heights for microwave-radio relay stations. Such tests provide information which cannot at present be reliably determined by any other means, and while the testing work is time consuming and involves some expense, the avoidance of serious problems which might otherwise be encountered after permanent construction is completed is believed to be more than adequate justification for use of this procedure on major radio-relay routes.

No Discussion

Computation of Impedance and Efficiency of Transmission Line with High Standing-Wave Ratio

W. W. MACALPINE
MEMBER AIEE

IT IS ordinarily desirable to operate a radio-frequency or other transmission line with standing-wave ratio as close to unity as possible. When the voltage standing-wave ratio (VSWR) is less than 10, means such as the Smith chart are available for making impedance computations with an accuracy sufficiently good for many purposes. However, in some cases it is desirable or unavoidable to operate the line with high standing-wave ratios. When the VSWR is around 100 the Smith chart is quite satisfactory for reactance but affords no degree of accuracy in reading resistance. This is due to the difficulty of reading even one significant figure on that scale near the periphery, and also to eccentricity in locating the movable arm or the dividers on the center point of the chart.

A simple method is shown for computing the resistance component when the impedance point lies close to the edge of the Smith chart. Attention is also given to precautions to be observed in computing normalized impedance. Similar procedures are shown for admittances. The effect of attenuation is taken into account.

In a related problem, the usual formulas for power and efficiency can lead to considerable error when the VSWR is high. A precise formula is developed herein for the power flowing at a location in a network, in terms involving the magnitude and angle of the reflection coefficient of the load with respect to the source. This is accurate both for lumped-constant networks and for those with distributed

---

Paper 53-192, recommended by the AIEE Special Communications Applications Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Southern District Meeting, Louisville, Ky., April 22-24, 1953. Manuscript submitted September 30, 1952; made available for printing March 2, 1953.

W. W. MACALPINE is with the Federal Telecommunication Laboratories, Inc., Nutley, N. J.

constants. It is in a form that makes it of particular value in connection with transmission lines in the transverse electromagnetic mode.

On the basis of the power formula, another is readily derived for the efficiency of power transfer along a section of transmission line. The complete formula is more involved than necessary for most applications but can be reduced to various simpler expressions according to the nature of the problem. One of these is the usual formula found for a line with low or medium standing-wave ratio. Another is much more accurate than the latter for cases where the standing-wave ratio is high. It is simple to use, involving only the ohmic resistances and the normalized reactances of the load and input impedances.

Symbols

The following list does not include some symbols that are defined where they occur in the text.

$A_0 = 8.686\alpha h$ = normal or matched attenuation of a length of line in decibels
B_0/G_0 = tangent of the phase angle of Y_0
$g+jb = Y/Y_0$
h = distance along line between two points in units of length
$r+jx = Z/Z_0$
R_0 = real part of the characteristic impedance
S = voltage standing-wave ratio
$Y_0 = 1/Z_0 = G_0(1+jB_0/G_0)$
$Z_0 = R_0(1-jB_0/G_0)$ = characteristic impedance in ohms
α = attenuation constant in nepers per unit length
β = phase constant in radians per unit length
δ = loss angle of dielectric, or power factor when $\delta < 0.1$
η = efficiency
$\theta = \beta h$
$\rho = |\rho|/2\psi$ = voltage reflection coefficient
ψ = electrical angle to a point on the line, measured toward the load from a voltage standing-wave maximum

Transformation of Impedance

The first step in solving a particular problem is to decide on the method that is simplest while satisfying the accuracy requirements. A typical case can be set up as follows with reference to Figure 1.

Given: The impedance $Z_1 = R_1 + jX_1$ at one end of the line, and the characteristic resistance R_0. Also α/β, B_0/G_0 and θ or h/λ, which are determined as indicated in the section on attenuation and line length.

Required: To find the impedance $Z_2 = R_2 + jX_2$ at the other end of the line and the efficiency.

GENERAL RULES

Determine the simple normalized impedance $(R_1 + jX_1)/R_0$ and set it up on the Smith chart. If the resistance component can be read accurately (say VSWR less than 10 or 20), it is ordinarily satisfactory and simple to use the chart for the entire computation.

If the VSWR is greater than 10 or 20 and the resistance component cannot be read on the Smith chart with the desired accuracy, the following method is recommended.

PROCEDURE

The normalized impedance should properly be computed by use of equation 4 and set up on the Smith chart. However, in most cases the simpler expression $(R_1 + jX_1)/R_0$ is sufficiently accurate for this first step. If working in admittances, use $(G_1 + jB_1)/G_0$. Determine the normalized impedance $r_2 + jx_2$ (or the admittance if required) at the other point of the line by the usual manipulation of the chart. Take account of attenuation by means of equations 9 and 10 and the remarks thereafter. The reactance $X_2 = R_0x_2$ thus found is sufficiently accurate for most purposes. Then subject to the following conditions, the resistance or conductance component can be computed by use of equation 1 and its modifications in equation 2 according to the given and required quantities.

$$R_2 = R_1\frac{1+x_2^2}{1+x_1^2} + R_0(1+x_2^2)\left[\frac{\alpha}{\beta}\theta + \frac{B_0}{G_0}\times \left(\frac{x_2}{1+x_2^2} - \frac{x_1}{1+x_1^2}\right)\right] \quad (1)$$

Observe that R is the ohmic resistance, while x is the normalized reactance. The attenuation of the line in nepers is shown as $(\alpha/\beta)\,\theta$ but can be expressed in other ways, as in equation 10. It is positive when point *2* is on the generator side of point *1*, and negative in the converse case.

Equation 1 is not as tedious to use as its length might indicate, for the terms $(1 + x_1^2)$ and $(1 + x_2^2)$ are each repeated two or three times. Also, in many problems $B_0/G_0 = \alpha/\beta$ affording further simplification.

CONDITIONS

Equations 1 and 2 yield results with accuracy of the order of 1 per cent subject to the following conditions.

1. Neither $1/S_1$ nor $1/S_2$ exceeds 0.17
2. $|B_0/G_0| < 0.1$
3. The normalized impedances r_1+jx_1 and r_2+jx_2 or admittances g_1+jb_1 and g_2+jb_2, whichever are used, must lie in the one-per-cent "permitted region" of Figure 2.
4. The line parameters and given impedance be known to 1-per-cent accuracy. In many problems, this last condition is not satisfied.

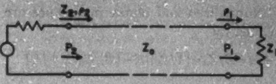


Figure 1. Transmission line of length h

VARIATIONS OF FORMULA

Equation 1 gives R_2 when impedance is given. When admittance is given or required, similar formulas can be written with terms shown in the following tabulation:

$$\left\|\begin{array}{lllllll} R_2 & R_1 & x_2^2 & x_1^2 & R_0 & x_2 & -x_1 \\ G_2 & G_1 & b_2^2 & b_1^2 & 1/R_0 & -b_2 & b_1 \\ R_2 & G_1R_0^2 & x_2^2 & b_1^2 & R_0 & x_2 & b_1 \\ G_2 & R_1/R_0^2 & b_2^2 & x_1^2 & 1/R_0 & -b_2 & -x_1 \end{array}\right\| \quad (2)$$

The top row shows the terms in equation 1. The second row is for G_2 in terms of G_1, etc. Replace each term in equation 1 with the second row term directly below the original term in the tabulation. Be sure to replace all the x_2^2 by b_2^2 and so on.

It may happen that R_2 is desired, but $r_2 + jx_2$ is not in the permitted region while $g_2 + jb_2$ is in it. Compute G_2 by equation 1 modified by equation 2. Instead of taking the reciprocal of $G_2 + jB_2$, use the equation

$$R_2 = R_0^2G_2|x_2/b_2| \quad (3)$$

where x_2 and b_2 are read on the Smith chart in the usual manner for converting impedances to admittances.

Similarly, when G_2 is to be found and $g_2 + jb_2$ is not permitted, compute R_2 and transpose equation 3 to find G_2. The same means is used to convert between R_1 and G_1 when necessary.

Normalized Impedance and Other Preliminary Considerations

NORMALIZED IMPEDANCE

The characteristic impedance and admittance are

$$Z_0 = R_0(1+jX_0/R_0) = R_0(1-jB_0/G_0)$$

$$Y_0 = G_0(1+jB_0/G_0)$$

The impedance and admittance looking toward the load at a point in the line are

$$Z = R+jX$$

$$Y = 1/Z = G+jB$$

Normalized and ohmic impedance are related by

$$r+jx=Z/Z_0=(1/R_0)[R-(B_0/G_0)X+jX] \quad (4)$$

$$R+jX=Z=R_0[r+(B_0/G_0)x+jx] \quad (5)$$

accurate to 1 per cent provided

$$R/|X|=G/|B|<0.1;\ r/|x|=g/|b|<0.1;\ |B_0/G_0|<0.1$$

The equations for admittances are similar, with each impedance term replaced by the corresponding admittance term and the sign before B_0/G_0 reversed.

Caution must be observed to retain the proper sign with X, B, x, or b in the formulas. For example if $X = -300$, it must not be entered in the equation simply as 300. Take special care in this for the terms with coefficient B_0/G_0.

The B_0/G_0 term in these equations can frequently be neglected leaving the customary simple equations for normalized impedance. This is especially so in rough preliminary design computations. The B_0/G_0 terms are most important when the load Q is high, or B_0/G_0 is relatively high such as in spiral lines and resonators and with ordinary lines used at the lower frequencies.

ATTENUATION AND LINE LENGTH

If the attenuation of the line is known in decibles per 100 feet, then

$$\frac{\alpha}{\beta}=\frac{0.115(\text{db}/100\ \text{ft})}{30.48}\,\frac{\lambda}{2\pi}\ \text{nepers per radian}$$

Wave length in meters for an ordinary line is

$$\lambda=v/f=300/f\sqrt{\epsilon}$$

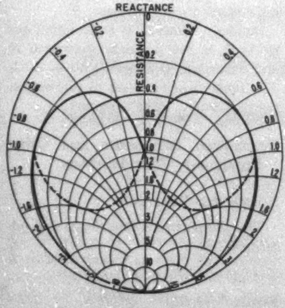


Figure 2. Permitted region for use of formula $S=(1+x^2)/r$. Area outside solid heart-shaped curve drawn on Smith chart is where equation is accurate to within 1 per cent. Area outside dashed curve is where reciprocal of $r+jx$ lies in permitted region, see Appendix I

When attenuation data are not available, compute

$$\alpha/\beta=Rv/2\omega R_0+\delta/2$$

The resistance of a concentric line or a balanced 2-conductor line with copper conductors is

$$R=\sqrt{f}(1/d+1/D)\times10^{-8}\ \text{ohms per foot}$$

where d and D are the diameters in inches of the active surfaces of the conductors and f is in megacycles.

The tangent of the phase angle of the characteristic admittance is

$$B_0/G_0=\alpha/\beta-\delta$$

The line length is

$$h/\lambda=hf\sqrt{\epsilon}/984\ \text{wave lengths}$$

$$\theta=2\pi h/\lambda=hf\sqrt{\epsilon}/156.5\ \text{radians}$$

where h is in feet and f in megacycles per second, while ϵ is the dielectric constant of the medium relative to air.

Derivation of Impedance Transformation Equation

The following equations are well known

$$1/S=(1-|\rho|)/(1+|\rho|)$$

$$|\rho_2|=|\rho_1|\epsilon^{-2\alpha h}=|\rho_1|\epsilon^{-0.230A_0}=|\rho_1|10^{-A_0/10} \quad (6)$$

With respect to a fictitious load with $|\rho|=1.00$

$$|\rho_1|=\epsilon^{-2\alpha h_1}$$

$$|\rho_2|=\epsilon^{-2\alpha h_2}$$

$$h=h_2-h_1$$

It follows that

$$1/S_2=\tanh[0.115A_0+\tanh^{-1}(1/S_1)] \quad (7)$$

Now utilize the results of Appendix I and Figure 2, and note that equation 7 reduces to equation 9 with an accuracy of 1 per cent when neither $1/S_1$ nor $1/S_2$ exceeds 0.17.

$$1/S_1=r_1/(1+x_1^2)=g_1/(1+b_1^2) \quad (8)$$

$$1/S_2=1/S_1+(\alpha/\beta)\theta \quad (9)$$

$$(\alpha/\beta)\theta=\alpha h=0.115A_0 \quad (10)$$

Finally

$$r_2=(1+x_2^2)/S_2$$

$$g_2=(1+b_2^2)/S_2 \quad (11)$$

When equations 8 to 11 are combined with equations 4 and 5 for normalized impedance, equation 1 results. The preceding equations can be used in a step-by-step method of computing impedance, but usually equation 1 provides a shorter computation.

Note the remarks under equation 1 on the sign of the attenuation. When attenuation is greater than 0.01 neper or about 0.1 decibel, it is desirable to take it into account in manipulating the Smith chart. This is done by means of equations 7 and 9, noting that the radial scale "resistance component" (graduated between 0 and 1.0) is identical with $1/S$. The "1-decibel steps" scale on the chart is too coarse for use in many problems.

Power and Efficiency

The net power flowing toward the load at a point on a transmission line Figure 1 is shown in Appendix III to be

$$P={}_fV^2G_0\left(1-|\rho|^2+2|\rho|\frac{B_0}{G_0}\sin 2\psi\right)\text{watts} \quad (12)$$

where ${}_fV$ is the rms amplitude of the voltage of the incident wave. The efficiency of power transfer between point *2* on the generator end and point *1* on the load end of a section of line is

$$\eta=\frac{P_1}{P_2}=\frac{1-|\rho_1|^2+2|\rho_1|(B_0/G_0)\sin 2\psi_1}{1-|\rho_2|^2+2|\rho_2|(B_0/G_0)\sin 2\psi_2}\epsilon^{-2\alpha h} \quad (13)$$

This equation can be simplified in many cases according to the conditions of the problem.

CONDITION A

$$S_1|B_0/G_0|\ll1$$

$$\eta=\frac{1-|\rho_1|^2}{1-|\rho_2|^2}\epsilon^{-2\alpha h}$$

$$=\frac{1/|\rho_1|-|\rho_1|}{1/|\rho_2|-|\rho_2|}=\frac{S_2-1/S_2}{S_1-1/S_1} \quad (14)$$

where $|\rho_2|$ and $|\rho_1|$ are related as in equation 6. The maximum per-cent error, due to the combined effects of numerator and denominator, is roughly

$$\pm100S_1B_0/G_0$$

This is the usual formula found for the efficiency of a radio-frequency transmission line. When the impedance can be computed accurately on the Smith chart, the VSWR being less than 10 or 20, equation 14 should ordinarily be used. The values of $|\rho_1|$ and $|\rho_2|$ can be read on the chart, but it is suggested that one be read and the other computed by equation 6, since they usually differ but little.

CONDITION B

$$S_2>>1$$

$$\eta=\frac{1/S_1+0.5(B_0/G_0)\sin 2\psi_1}{1/S_2+0.5(B_0/G_0)\sin 2\psi_2} \quad (15)$$

The per-cent error is roughly

$$100/(S_2^2-1)-100/(S_1^2-1)$$

which is

<1 per cent for $S_2>10$ and <10 per cent for $S_2>3.3$

CONDITION C

$S_2>>1$ and $S_1|B_0/G_0|<<1$

$$\eta=S_2/S_1=1/(1+S_1\alpha h)=1/(1+0.115A_0S_1) \quad (16)$$

The error is the algebraic sum of those for conditions *A* and *B*. In the expressions involving αh and A_0, this error is increased slightly by an approximation $\tanh x\approx x$ in the use of equation 7.

CONDITION D

$S>>1, |B_0/G_0|<<1$ and $r+jx$ or $g+jb$ in the "permitted region" of Figure 2

$$\eta=\frac{R_1}{R_2}\left(\frac{1+x_2^2}{1+x_1^2}\right)=\frac{G_1}{G_2}\left(\frac{1+b_2^2}{1+b_1^2}\right)$$
$$=\frac{R_1}{R_0^2G_2}\left(\frac{1+b_2^2}{1+x_1^2}\right)=\frac{R_0^2G_1}{R_2}\left(\frac{1+x_2^2}{1+b_1^2}\right) \quad (17)$$

The accuracy of equation 17 is comparable to that of equations 1 and 2 under the conditions listed for those equations.

R_1 and G_1 are the resistive and conductive components of the ohmic impedance and admittance looking toward the load at point *1*, and similarly for R_2 and G_2. On the other hand, *x* and *b* are normalized values. The real part of the characteristic impedance is R_0.

In problems where it is desirable to use equations 1 and 2 in computing impedance, equation 17 is generally the simplest one to use for efficiency and the most accurate for numerical computations. The latter equation is identical with the first term on the right-hand side of equation 1 or 2 divided by R_2 or G_2 as the case may be.

The exponential does not appear explicitly in several of the equations 14 to 17. This is not an approximation for small values of attenuation, but the exponential has been absorbed into the body of the equation.

The stated errors for the equations are expressed as a percentage of η, obviously not in addition to η. Furthermore, the impedance and line parameters often will be known inaccurately and can introduce considerable additional error into a numerical computation.

Appendix I. Equations for Boundary of Permitted Region

For 1-per-cent or better accuracy in equations 8 and 11 it is now shown that the impedance point must be in the permitted region of Figure 2 for which

$$|\cot\psi|<0.1S^2/\sqrt{S^2-1} \quad (18)$$

For 10-per-cent accuracy, substitute the coefficient 0.32 for 0.1. When working in admittances, use $\tan\psi$ in place of $\cot\psi$. This gives the same boundary line in Figure 2.

Good formulas for checking rapidly the validity of equations 8 and 11 for 1-per-cent accuracy are

$$r<0.1|x+1/x|$$
$$g<0.1|b+1/b| \quad (19)$$

provided $|x|>0.3$ or $|b|>0.3$ as the case may be. The accuracy condition can be checked visually on the Smith chart, where the circles of constant *x* must not deviate from the radial direction by more than about 10 degrees at the impedance point in question.

Expression 18 is derived as follows. The complex reflection coefficient can be written

$$\rho=[(S-1)/(S+1)](\cos 2\psi+j\sin 2\psi)$$

The normalized impedance is given by

$$r+jx=(1+\rho)/(1-\rho)=(1+jS\cot\psi)/(S+j\cot\psi)$$

The expression in *S* and ψ can be demonstrated most readily by substituting the preceding value for ρ and equating as a trigonometrical identity. It is an exact formula. Although it corresponds to an impedance formula with attenuation neglected, the fact that attenuation does not appear does not constitute an approximation. The quantities *S* and ψ are parameters that completely determine the reflection coefficient at the point where the impedance is $r+jx$.

Equation 8 can be derived from the last formula

$$r/(1+x^2)=(1/S)(S^4+S^2\cot^2\psi)/(S^4+\cot^2\psi)=(1+a)/S$$

The correction term *a* is always positive.

$$a=(S^2-1)\cot^2\psi/(S^4+\cot^2\psi)\approx[(S^2-1)/S^4]\cot^2\psi \quad (18A)$$

For 1-per-cent error in equation 8, $a=0.01$ and the expression 18 results.

Appendix II. Effect of B_0/G_0 Terms in Impedance Equations

For this discussion, it is convenient to modify equation 1 by the use of the apparent or approximate standing-wave ratio S' and by a result found in Appendix IV.

$$1/S'=R/R_0(1+x^2)\approx G/G_0(1+b^2)$$

$$\sin 2\psi\approx 2x/(1+x^2)\approx -2b/(1+b^2)$$

Then equations 1 and 2 become

$$\frac{1}{S_2'}=\frac{1}{S_1'}+\frac{\alpha}{\beta}\theta+\frac{1}{2}\frac{B_0}{G_0}(\sin 2\psi_2-\sin 2\psi_1) \quad (20)$$

The angle of the reflection coefficient 2ψ can be read on the Smith chart. If working with admittances, read the negative of the angle indicated on the periphery for practical purposes in computing $\sin 2\psi$.

The principal point in question is the effect of the B_0/G_0 terms in the over-all computation and whether they can be neglected. The worst case is when $B_0/G_0=\pm\alpha/\beta$. Then we are interested in the relative value of the quantity

$$(1/2)(\sin 2\psi_2-\sin 2\psi_1) \quad (21)$$

compared to the line length θ. If θ is centered about a voltage maximum or minimum, the quantity 21 is greatest with maximum value of unity. The other extreme is for θ centered midway between these points, when quantity 21 is zero. For small values of θ up to about a radian, the quantity can range between almost doubling the effect of θ down to almost annulling it.

Even with low VSWR, the quantity 21 has the same effect as described on θ. However, when θ is small, it does not have appreciable effect on $1/S$ or the impedance relationship in this case. When θ is large enough so attenuation has appreciable effect on impedances at low VSWR, the quantity 21 is negligible in comparison.

Then the B_0/G_0 terms can be neglected when the standing-wave ratio is low. For higher VSWR, use or omit the B_0/G_0 terms in equations 1, 2, 4, and 20 depending on a mental comparison of the magnitude of quantity 21 in relation to θ in radians.

Appendix III. Derivation of Power Equation

The power flowing at a point in an a-c circuit is given by

$$P=(\text{Real})VI^* \quad (22)$$

This is the real part of the product of the rms complex sinusoid voltage by the conjugate of the corresponding current.

Equation 12 is easily found when the following expressions are substituted in this equation. At a point in a transmission line

$$V={}_fV(1+\rho)$$

$$I={}_fVY_0(1-\rho)$$

$$I^*={}_fV^*Y_0^*(1-\rho^*)$$

$$Y_0^*=G_0(1-jB_0/G_0)$$

$$\rho=(Z-Z_0)/(Z+Z_0)=|\rho|\epsilon^{j2\psi}$$

$$\rho^*=|\rho|\epsilon^{-j2\psi}$$

The power equation 12 can also be demonstrated by a related method. At a point on a line the power flow is given by $P=V^2G$ where *V* is the rms voltage across the line and *G* is the conductive component of the admittance looking toward the load. When this is expressed in terms of reflection coefficient, etc., equation 12 results.

The equation can also be demonstrated by integrating the power loss in the series re-

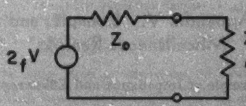


Figure 3. Generator and load

sistance and shunt conductance of the line, between a known load and any other point on the line.

An example and further remarks on the power in a line are given at the end of Appendix V.

Appendix IV. Derivation of Efficiency Equations

The complete equation 13 follows immediately from equation 12. In deriving equation 15 there is used equation 6 and

$$1-|\rho|^2=(1+|\rho|)^2/S=(4|\rho|/S)(1+a') \qquad (23)$$

$$a'=1/(S^2-1)$$

The expressions 17 are derived from equation 13 by way of equation 15 with the following steps. From equations 8, 18(A), and 23

$$(1-|\rho|^2)/4|\rho|=(1+a')/S$$
$$=(1+a'-a)r/(1+x^2)$$

$$a'-a\approx(1-\cot^2\psi)/S^2$$

By a deviation similar to that of Appendix I

$$0.5\sin 2\psi=x/(1+x^2)(1-a'')$$

$$a''=[S^2-(S^2-2)\cot^2\psi]/(S^4+\cot^2\psi)$$
$$\approx(1-\cot^2\psi)/S^2$$

This is the same as the approximate equation for $a'-a$ already found. Then

$$1-|\rho|^2+2|\rho|(B_0/G_0)\sin 2\psi$$
$$=(1+a'')4|\rho|[r+(B_0/G_0)x]/(1+x^2)$$
$$=(1+a'')4|\rho|R/R_0(1+x^2)$$

with equation 5 used in the last step. A similar equation is found for admittances. Equation 17 follows readily.

The physical significance of equation 17 is as follows. Subject to the stated conditions, there holds approximately

$$x=\cot\psi$$

$$I=I_{\max}\sin\psi$$

where $I_{\max}$ is the current standing-wave maximum. The power at any point is

$$P=I^2R$$

where R is the resistive component of the ohmic impedance looking toward the load at the point. Equation 17 follows, since

$$I^2=I_{\max}^2\sin^2\psi=I_{\max}^2/(1+\cot^2\psi)$$
$$=I_{\max}^2/(1+x^2)$$

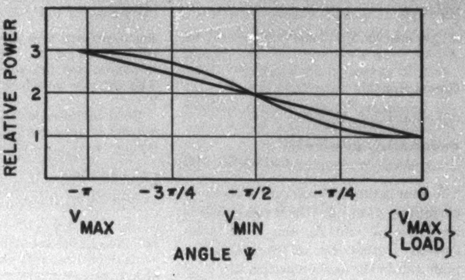
The value of $I_{\max}$ is constant within practical limits for values of S greater than those listed in the statement of errors under the section entitled "Condition D."

Appendix V. Examples and Miscellaneous Remarks

Example of Impedance and Efficiency

A load of $5-j1{,}000$ ohms at 2.0 megacycles is fed by a 60-foot length of *RG-17/U* cable. What is the input impedance and efficiency?

Figure 4. Power flow along line. Straight line is computed power neglecting B_0/G_0 term in equation 12. Wavy line is according to complete equation for case of negligible dielectric loss



The attenuation of the cable at 2.0 megacycles is found to be about 0.094 decibel per 100 feet. The loss angle of the polyethylene dielectric is less than 0.0002, and the dielectric constant is 2.26. The given conditions can be summarized as follows, making use of equations in the section on attenuation and line length.

$$R_1+jX_1=5-j1{,}000\text{ohms}$$

$$R_0=52 \text{ ohms}$$

$$\alpha/\beta=B_0/G_0=5.6\times10^{-3}$$

$$\theta=1.15\text{radians}(66°)$$

$$h/\lambda=0.183$$

Then $x_1=-1{,}000/52=-19.2$. By the Smith chart $x_2=-0.384$ and so $X_2=-20.0$ ohms. The resistance is found by equation 1 where $1+x_2^2=1.148$ and $1+x_1^2=370$

$$R_2=5\times\frac{1.148}{370}+52\times1.148\times5.6\times10^{-3}\times$$
$$\left(1.15-\frac{0.384}{1.148}+\frac{19.2}{370}\right)$$
$$=(15.5+290)10^{-3}=0.305 \text{ ohm}$$

Efficiency, by equation 17 is

$$\eta=15.5\times10^{-3}/0.305=0.051 \quad (5.1 \text{ per cent})$$

Reflection Coefficient Greater than Unity

Suppose a loss line with characteristic impedance $Z_0=R_0(1-j0.1)$ is terminated by a lossless load with impedance $Z=jR_0$. Then the normalized load impedance is found to be

$$Z/Z_0=-0.1+j1.0$$

The negative normalized resistance can be visualized if a phasor diagram of the impedances is drawn. It has no implication of negative ohmic resistance.

The reflection coefficient, see Appendix III, is

$$\rho=1.11\underline{/90.3°}$$

Now it is found that there is interchange of power between the incident and reflected waves, but the sum of all the power components is zero at the load.

$$(\text{Real})_f V_f I^*=1.00$$

$$(\text{Real})_r V_r I^*=-1.22$$

$$(\text{Real})_f V_r I^*=0.11$$

$$(\text{Real})_r V_f I^*=0.11$$

The popular conception that the magnitude of the reflection coefficient cannot exceed unity is correct when either the source or load impedance is a pure resistance. With a transmission line, this is usually justified. In fact, when Z_0 includes an appreciable reactive component, it follows inherently that the attenuation per radian is large. Then even if the reflected coefficient exceeds unity at the load, its magnitude is reduced rapidly along the line and soon is less than unity.

Power in Lumped-Element Network

It may be observed that the derivation in Appendix III of equation 12 shows it to be equally valid for either lumped or distributed networks. For the lumped case shown in Figure 3, Z_0 is the generator impedance, Z the load, the $_fV$ is half the open-circuit generator voltage. This equation is not required in most lumped-element network applications since other equations are usually simpler to apply.

With lumped-element networks, the reflection coefficient can be very large. Take for example a load $Z=R+jX$ connected to a generator impedance $Z_0=\boldsymbol{Z}^*=R-jX$. The reflection coefficient is $\rho=jX/R=\pm jQ$. If the open-circuit generator voltage is E, then the power into the load is $P=E^2/4R$ by inspection. It is interesting to compute the power by equation 12 which yields the same result.

One use for the idea of reflection coefficient with lumped element networks is in connection with the application of the principle of superposition. For instance, the load impedance may be replaced by one equal to the source impedance in series with a voltage equal to the actual open-circuit voltage of the source multiplied by the complex reflection coefficient.

Example of Power Flow Along Line

A practical interpretation is given of the power equation 12, showing the net power flowing toward the load at various points along a line with high attenuation.

Let $_fV_a$ and ρ_a be the values of incident rms voltage and reflection coefficient at a voltage maximum. Then at any other point

$$_fV = {_fV_a}\epsilon^{-\psi\alpha/\beta}$$

$$|\rho| = |\rho_a|\epsilon^{2\psi\alpha/\beta}$$

Now equation 12 can be written

$$P = {_fV_a}^2G_0[\epsilon^{-2\psi\alpha/\beta} - |\rho_a|^2\epsilon^{2\psi\alpha/\beta} + 2|\rho_a|(B_0/G_0)\sin 2\psi]$$

Suppose $\alpha/\beta = B_0/G_0 = 0.0318$ in which case the dielectric loss is negligible. Let $|\rho_a| = 0.9$ and find the power for various values of ψ between 0 and $-\pi$ radians.

The results are plotted in Figure 4, where arbitrarily $P = 1.00$ at $\psi = 0$, so ${_fV_a}^2G_0 = 5.25$. As a sample computation, let $\psi = -\pi/4$. Then

$$-2\psi\alpha/\beta = 0.050$$
$$P = 5.25[1.051 - (0.810/1.051) - 2\times 0.9\times 0.0318] = 1.18$$

The curve of power versus position on line in Figure 4 shows that the slope is almost zero at the current minimum and greatest at the current maximum. This agrees with the assumption of negligible dielectric loss, or $B_0/G_0 = \alpha/\beta$. On the other hand, if the conductor loss had been negligible but the dielectric loss appreciable, there would have resulted $B_0/G_0 = -\alpha/\beta$. Then the lobes of the sinusoidal curve of power would have been on the other side of the average power line (that is, where the B_0/G_0 term is neglected). The slope would be greatest at the voltage maximum and almost zero at the voltage minimum which is logical.

The solution of the differential equation of propagation shows uniform voltage and current attenuation of the incident wave and of the reflected wave as they flow in their respective directions. This condition is used in the derivation of equations 12 and 13. However, as shown in the example, this leads to the power dissipation varying nonuniformly along the line. The discrepancy is resolved by the transfer of power between the incident and reflected waves as shown earlier in this appendix.

Reference

1. Reference Data for Radio Engineers. Federal Telephone and Radio Corporation, New York, N. Y., third edition, 1949, chapter 16.

◆

No Discussion

The XY Toll Ticketing System

H. L. FOOTE
ASSOCIATE MEMBER AIEE

IN RECENT years independent telephone companies have become acutely aware of steadily decreasing net profits on short-distance toll-call operation. The recording, for billing purposes, of certain information concerning the toll call has required the assistance of an operator. For this reason, manual toll switching has been retained in spite of the fact that mechanical systems for performing the actual switching functions have been available for many years. Fixed toll charges and rising labor costs have combined to make it extremely difficult for the operating companies to realize a profit on this phase of their business.

One solution to this problem is the granting of free service on these loss-producing toll lines. While this would prevent further losses, resulting increased traffic would create a demand for additional expensive plant equipment and personnel which in turn would produce no revenue.

A second solution is the reduction of the number of people involved in producing the toll billing records known as "toll tickets," by the installation of automatic toll switching equipment incorporating automatic ticketing means. Several attempts have been made to provide such equipment and two basic types have emerged.

One, the verifying type, requires the calling subscriber to dial his own directory number before dialing the number of the desired station. The other, considerably more complex and expensive, is the identifying type wherein the calling station number is automatically found and recorded without effort on the part of the subscriber.[1] The system described herein is of the verifying type.

Any toll ticketing system, manual or automatic, must accomplish the following three things:

1. The information concerning the toll call must be recorded in some manner.
2. The recorded information must be translated and interpreted for use by the ticketing device.
3. The information must finally be presented in a readily usable form.

Recording the Information

Since the primary objective of any toll ticketing system is the reduction of manual labor in producing the toll tickets, the most desirable system would be the one which would reduce the labor factor to zero. In order to achieve such a goal, the designer is confronted with the need for an inexhaustible medium on which to store temporarily the billing information until that information is printed as a toll ticket. Several such storage devices are available, only one of which is economically sound.

The Recorder

The unique magnetic tape recorder shown in Figure 1 forms the heart of the XY toll ticketing system as the intermediate storage device. The tape in the magazine can record information concerning 100 average toll calls before it requires automatic interpreting equipment to print the tickets, simultaneously preparing itself for reuse. One recorder is permanently associated with each toll trunk, operating completely unattended in all of its functions. The recorder records the calling and called subscribers' directory numbers, the duration of the call (elapsed time), the time of day, and the date.

Figure 2 shows the XY switch manufactured by Stromberg-Carlson Company. This is a universal switch for use in any circuit in a step-by-step dial telephone system. Comparison with Figure 1 will show the similarity between the recorder and the switch. Great savings in tooling and manufacturing costs were achieved by using as many identical parts as possible in the two devices. Both are built on the same plate, use two of the same magnets, mount in the same way, and are jacked into the associated trunk relay equipment by means of identical cord and plug assemblies. Thus, the entire design of the XY toll ticketing equipment can be harmonious with standard XY switching equipment.

Paper 53-184, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Southern District Meeting, Louisville, Ky., April 22–24, 1953. Manuscript submitted January 21, 1953; made available for printing March 11, 1953.

H. L. Foote is with the Stromberg-Carlson Company, Rochester, N. Y.

The author wishes to acknowledge the contributions of F. A. Morris, J. D. Confeld, and H. S. Gleason who should be credited with the major portion of the circuit and mechanical design of the XY toll ticketing system.

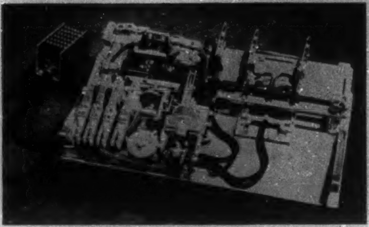


Figure 1 (left). Trunk recorder-reproducer used in the XY toll ticketing system

Figure 2 (right). The Stromberg-Carlson XY switch

The recording is made on standard 1/4 inch magnetic tape. Approximately 30 feet of tape are contained in the magazine in the form of an endless loop. The tape is transported by a capstan which can be driven by either of two methods. For recording, the tape is advanced step by step in 0.057-inch increments by means of a ratchet drive. For playback, a clutch connects the capstan to a common motor shaft and the tape is driven at a constant rate of about 3 inches per second. The tape passes over a twin-track recording head and a double-width erase head. Felt-faced pads apply the necessary pressure to assure proper contact between the tape and the heads. Three contact springs, sensing two conductive areas applied to the back side of the tape, control the cycling of the recording and playback process.

THE TAPE HEADS

Figure 3 shows the twin-track record-playback head before and after it has been cast in resin. Each head consists of a single 0.112-inch wide lamination of molybdenum permalloy with a 1-mil gap spacer. The coils are 2,700 turns of no. 46 wire wound on a bobbin. The various parts are mutually supporting and interlocking. The erase head is built on the same principle, the lamination being 0.270 inch wide, and its coil is 50 turns of no. 27 wire.

The upper half of the twin-track head is energized by each dial impulse and makes a magnetic impression 0.002-inch wide on the upper half of the tape. The lower half of the head is energized upon release of the interdigit relay in the trunk circuit which also affects the advance of the tape by a single step. These interdigit pulses serve to separate the several groups of pulses derived from the dialed digits. The two series of pulses are known as "mark" and "space" pulses respectively. To indicate the end of a call, mark and space pulses are recorded simultaneously when the trunk is released.

BLOCK DIAGRAM

Figure 4 shows a block diagram of the XY toll ticketing system, while Figure 5 shows the equipment itself, except for the trunk circuit. The digits of the subscriber's own number are used for two purposes. They are recorded in order to bill the call and they also set up a connection to verify the accuracy of the number dialed. If the call is verified, the verifying circuits are dropped off, and subsequent pulses (of the called number) are passed over the trunk to establish the connection to the distant subscriber. As soon as the called subscriber answers (answering supervision), the trunk transfers the recorder input to a common time impulse circuit, the minute pulser, transmitting one impulse per minute to the recorder during the conversation. When the calling subscriber hangs up (on-hook supervision), the trunk drops the linkage and transfers the input of the recorder to a common clock-calendar. Digital representation of the time of day and date is automatically pulsed into the recorder. Upon completion of this operation, the trunk releases and causes an end-of-call signal to be recorded. The trunk and its associated recorder are then free to accept another call.

If the calling subscriber dials other than his own number the call will not be verified and subsequent pulses are neither recorded nor repeated to the distant office. This condition is indicated to the subscriber by means of a suitable tone. Upon receiving on-hook supervision the trunk will merely enter an end-of-call signal and release. Similarly, in the case of "don't answer" and busy line calls, the trunk will enter an end-of-call signal—omitting the time and date information—and release. The presence of the date and time of day information is the criterion for the playback equipment to print a valid ticket. Abnormal calls are thus ignored by the printing equipment.

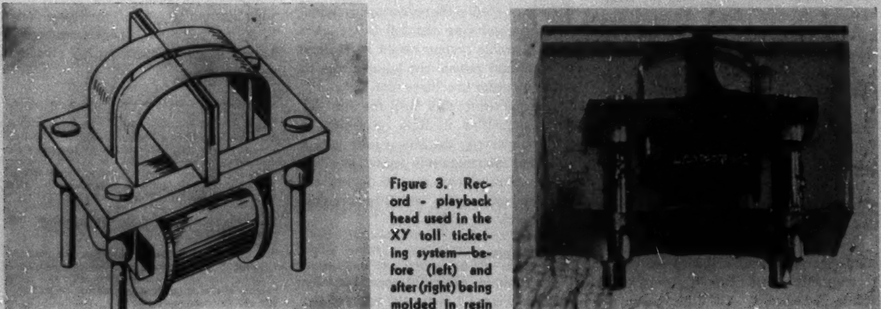


Figure 3. Record - playback head used in the XY toll ticketing system—before (left) and after (right) being molded in resin

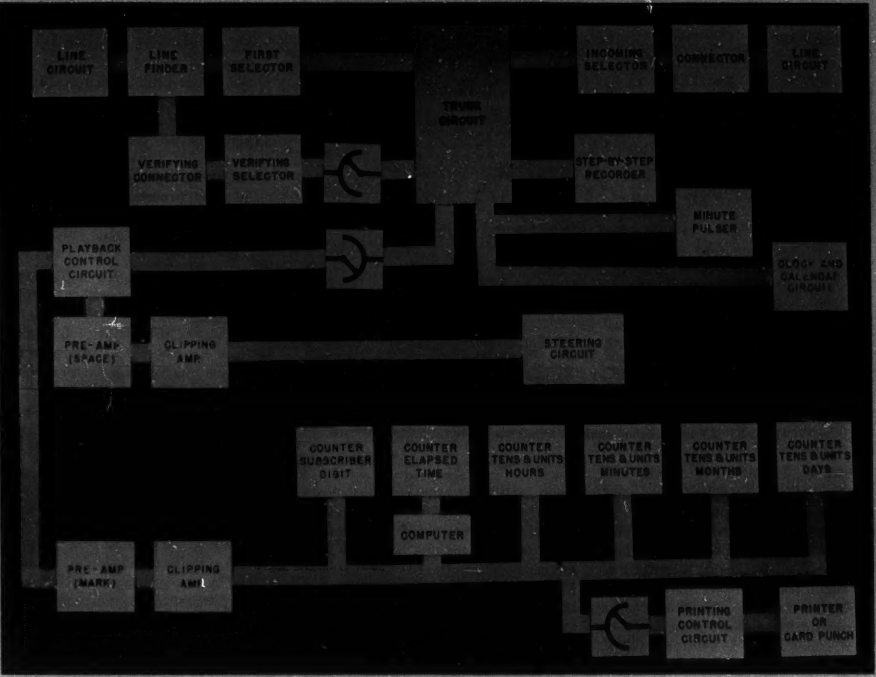

LINE CIRCUIT
LINE FINDER
FIRST SELECTOR
TRUNK CIRCUIT
INCOMING SELECTOR
CONNECTOR
LINE CIRCUIT
VERIFYING CONNECTOR
VERIFYING SELECTOR
STEP-BY-STEP RECORDER
MINUTE PULSER
PLAYBACK CONTROL CIRCUIT
CLOCK AND CALENDAR CIRCUIT
PRE-AMP (SPACE)
CLIPPING AMP
STEERING CIRCUIT
COUNTER SUBSCRIBER DIGIT
COUNTER ELAPSED TIME
COUNTER TENS & UNITS HOURS
COUNTER TENS & UNITS MINUTES
COUNTER TENS & UNITS MONTHS
COUNTER TENS & UNITS DAYS
COMPUTER
PRE-AMP (MARK)
CLIPPING AMP
PRINTING CONTROL CIRCUIT
PRINTER OR CARD PUNCH

Figure 4. Block diagram of the XY toll ticketing system

INTERPRETATION AND PRESENTATION

Common playback equipment attaches itself to each trunk recorder according to a prearranged schedule to permit common interpreting equipment to analyze the information recorded on the magnetic tape recorder. The interpreting equipment operates a ticket printer, common to all recorders, which prints a standard size toll ticket and automatically stacks these tickets, printing up to 10,000 per day. If complete automatic billing is desired, the information can be directed into a card-punching machine instead of the printer. In this case, 15,000 calls can be ticketed per day, since the card punch will operate somewhat faster than the printer.

PLAYBACK CONTROL CIRCUIT

Due to the large storage capacity of the trunk recorders, it is necessary and desirable to interpret the stored information only periodically, the period being determined by the toll traffic. Presumably, this would be once a day during a time of light traffic. The common clock-calendar unit is so wired that it will start playback at a predetermined time. The playback control circuit consists of a rotary switch and relays which function on a common equipment basis to augment the trunk during the playback cycle. The rotary switch acts as a trunk finder, stepping until the first idle trunk is found and associating the common control equipment upon seizing the trunk.

In the event that unusually heavy traffic conditions should cause a tape to become completely filled before the normal playback period, the normal schedule is ignored and the services of the playback control circuit requested at once. The rotary switch will advance directly to the full trunk and begin ticketing the calls from the recorder.

The "full" tape condition is sensed by the two conductive areas (foils) on the back of the tape. Should a subscriber seize the trunk and recorder just before the first foil is sensed, the recording of the data concerning the call will carry the tape past this point. Upon receiving on-hook supervision, the trunk immediately busies itself out and calls for playback as described previously. A "safety zone" has been provided to take care of long conversations in this area. Should the subscriber pass the first foil and the conversation continue for 99 minutes, the second foil will be sensed—indicating an absolute "deadline." However, a super-safety area has been provided so that when the second foil gives a forced disconnect, there still remains sufficient tape on which to record the date and time of day. If desired, the call can be continued by a manual operator rather than forcing disconnect.

PRINTING CONTROL CIRCUIT

The clutch of the recorder of the seized trunk is energized to move the tape to its "home" position. This is determined by the foil on the back of the tape. Control of the clutch is then transferred to the printing control circuit and the outputs of the twin record-playback head are con-

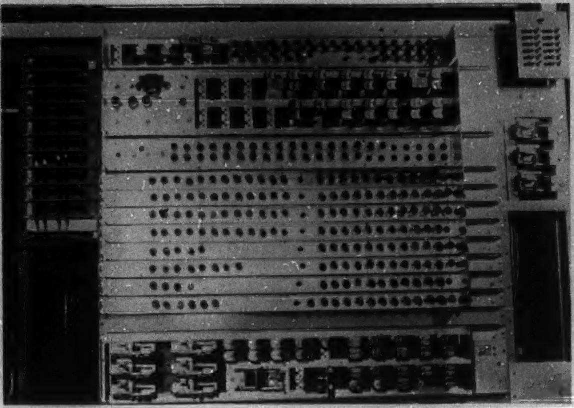


Figure 5. XY toll ticketing equipment (less trunks and recorders)

nected to the mark and space preamplifiers. The recorder remains under the control of the printing control circuit until the tape has been completely played back and erased, whereupon the playback control circuit takes over once again. Before the trunk is released, a single reset signal is recorded for purposes to be discussed later.

Upon releasing the trunk, the rotary switch advances and associates the playback control circuit with the next idle trunk to repeat the playback process. After completing a search through all of the trunks, the rotary switch will reach home and the playback control circuit will normally become inactive until the next scheduled start. After being amplified in the preamplifiers, the mark and space pulses are fed to the clipping amplifiers which further amplify the signals to a level suitable for driving the counting chains. The clipping amplifiers also act as marginal devices and will not transmit low-level signals resulting from interference. The signal pulses are shaped and brought to a uniform size by these amplifiers which have a very large range. Pulses having as little as 30 per cent of "standard" amplitude are transmitted and restored to full amplitude. The cutoff below that point is very rapid so that pulses only slightly less are completely eliminated. Pulses may be many times larger than standard and may still be restored.

STEERING CONTROL CIRCUIT

The function of the steering control circuit is to energize the proper circuits to count and store the recorded digits. When the steering control circuit is primed, it enables the first counter to count and store the first digit. Upon receiving a space pulse, the steering chain will be advanced to the next step, enabling the second counter to count and store the second digit, and so forth. The steering chain will have as many stages as there are total digits of billing information.

COUNTERS

With a playback tape speed of about 3 inches per second, the stored pulses are read off the tape at about 52 per second. Relay counting chains for counting and storing at this rate are extremely complex and expensive. Cold-cathode counting chains, on the other hand, can count very comfortably at this rate. One counting chain is used per digit of billing information, but the several chains vary somewhat in number of counting stages depending upon the maximum number of pulses expected. Those used for subscriber digits must count to ten, whereas those counting tens days, for instance, need count only to three. Where a large number of pulses must be converted into a 2-digit decimal number (for example, the elapsed time can have as many as 99 pulses) there is no single symbol for the quantity 99 and it must be represented as two 9's. This is accomplished by connecting the ends of a counting chain together to form a "ring." This ring will represent the units digit of elapsed time. It will continue to count around, repeating as many times as necessary. However, for every time the units ring returns to zero, one pulse will be fed into an associated chain to represent the tens digit of the elapsed time. Thus, if 45 pulses, representing 45 minutes of elapsed time, are fed into the counter, the units ring will return to zero four times, thus stepping the tens chain to four and will stop on the fifth tube in the ring representing units digit five. This automatically converts a continuous series of pulses into two decimal digits.

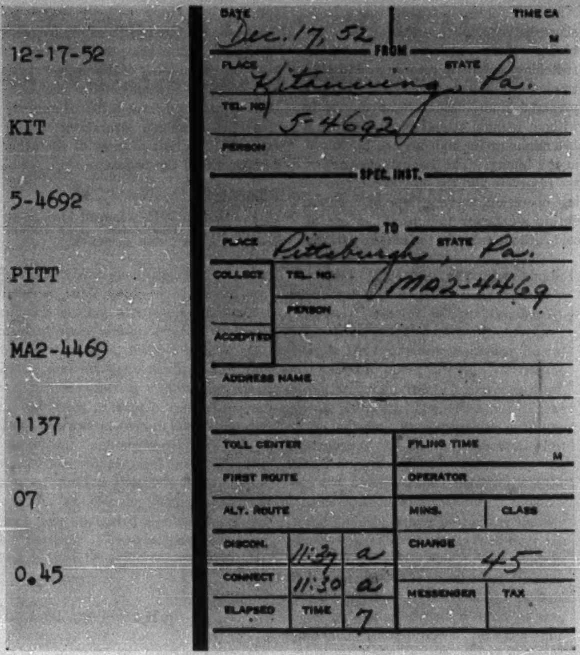
INTERPRETATION

The information stored in the several counters must be transferred to the ticket printer a digit at a time. By connecting the anodes of the counter tubes to the banks of a rotary switch it is possible to determine which tube is fired in each of the counters. Note that the information thus stored may be read, and the ticket printed, in any order. A printed ticket is shown with a manually written ticket in Figure 5.

Due to the requirements of the ticket printer it is necessary to translate the decimal information into a permutation indicative of the numeral or letter to be printed. This translation is accomplished by a group of relays which are connected by the rotary switch to the anodes of the counter tubes. The ticket printer employs a 7-unit code so arranged that the operation of any single translating relay will result in printing a numeral. Combinations of two relays will result in printing a letter. The ticket printer and the printing control circuit are interlocked by the simple expedient of advancing the rotary switch by means of a pulse transmitted from a cam in the ticket printer.

Some of the information on the toll tickets is determined by means other than recorded data. The names of both the originating and called offices are determined by the operation of relays under the control of the playback control circuit which put the proper letter codes on the proper terminals of the rotary switch banks. Likewise, the digits representing the year are strapped on the proper bank terminals. Information needed by the printer such as carriage return, line space, ejection, etc. is strapped in.

The printing control circuit acts to correlate the functions of the playback control, steering, counters, and printing control circuit. Included in the printing control are the "destination code" relays which determine what trunk is being ticketed and thus control what office letters are printed on the ticket. These same relays are also used to adjust the



12-17-52

KIT

5-4692

PITT

MA2-4469

1137

07

0.45

DATE Dec. 17, 52
TIME CA M
FROM
PLACE Kittanning STATE Pa.
TEL. NO. 5-4692
PERSON
SPEC. INST.
TO
PLACE Pittsburgh STATE Pa.
COLLECT
TEL. NO. MA2-4469
PERSON
ACCEPTED
ADDRESS NAME
TOLL CENTER
FILING TIME M
FIRST ROUTE
OPERATOR
ALT. ROUTE
MINS.
CLASS
DISCON. 11:37 a
CHARGE 45
CONNECT 11:30 a
MESSENGER
TAX
ELAPSED TIME 7

Figure 6. A toll ticket produced by the XY system compared with a manually written toll ticket

computer for the proper base time and base rate.

The printing and playback control circuits are carefully interlocked so that playback cannot proceed unless the counters are in a receptive condition nor can control by the printing control be relinquished until the information is completely interpreted. To make certain that the counters do not contain any information left over from previous ticketing or other interference the playback control transmits a reset signal each time a new trunk is seized. As an additional precaution a reset signal is also recorded at the beginning of each tape as mentioned earlier. The reset signal is simply an end-of-call recording, which, when not preceded by normal ticketing information, serves to restore all of the counting chains to the starting condition. After the two safety resets, the counters receive and store the call information, storing all information about one call before any other action takes place.

The Computer

The computer makes use of the elapsed time and destination code information to derive the charge for the call. It too uses cold-cathode counting chains to calculate the charge. Essentially, there are two chains and two rings. One chain counts off the minimum time for the particular call, the number of stages in the chain being adjusted by the destination code relays. If the call is based on a 3-minute period, for example, there will be three stages in the chain.

The first ring counts off the number of 5-cent increments (since all short-haul charges increase at the rate of 5 cents per minute), the second ring counts the ten cents, and the second chain counts the dollars.

The minimum charge is primed into these counters by means of the destination code relays. If the call is based on 15 cents minimum charge, the 5-cent tube and the number 1 tens cents tube will be primed and the charging will start at that point.

The elapsed time pulses are fed into the computer which at first counts them in the minimum-time chain. When the end of this chain is reached, the remaining pulses are diverted into the first money ring which will add 5 cents for each additional minute—the total being kept by the remaining two counters. The calculated charge is read by wiring the anodes of the counters to the rotary switch just as the numerical information in the other counters.

Ticket Printer

The basic mechanism of the ticket printer shown in Figure 6 is an electric typewriter to which the Commercial Controls Company adds what is known as a "translator" to make it respond to electrical control. Essentially, the translator is much the same as the selecting mechanism in a teletype machine.

For the XY toll ticketing system the ticket printer is further modified to the extent of adding a device for feeding roll paper into it and converting the continuous roll into individual tickets neatly stacked.

The paper rolls are sufficient for 10,000 tickets, the paper being standard toll-ticket width. It is fed to the platen through guide channels, and after passing around the platen and being printed, it passes through another guide channel to a solenoid-operated chopping knife. After being cut off, the ticket is ejected into a stacking hopper.

Power Requirements

The only power required for recording the toll-call information is the normal office battery at 48 volts d-c. For play-

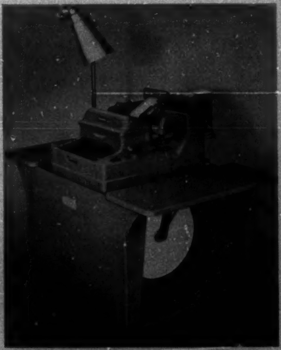


Figure 7. The ticket printer used in the XY system

ing back and ticketing, 110-volt a-c power is needed. The anode voltage for the cold-cathode tubes is derived from a simple 500-milliampere 150-volt d-c power supply.

Conclusions

The XY toll ticketing system meets the needs of the small and medium size telephone companies in their endeavor to solve the problem of decreasing profits in short-distance toll operation. By combining three basic arts—mechanical switching, magnetic recording, and electronics—the tickets for the toll calls are prepared in any desired form completely without manual attention, thus reducing to a minimum the high labor cost involved in toll billing. The 100-call capacity of the recorders and the 48-volt d-c recording operation prevent loss of revenue during a-c power failures. The use of cold-cathode tubes in the majority of the electronic circuits precludes the frequent changing of tubes since their life depends only on the number of on-off operations to which they are subjected. The *RCA-5823* tubes used have a life of approximately 40 million operations—a life equivalent to that of most of the other components of the system.

References

1. **Fundamentals of the Automatic Message Accounting System, John Meszar.** *AIEE Transactions*, **vol. 69, part I, 1950, pp. 255–69.**

No Discussion

A Capacitively Coupled Field Mapper for 2-Dimensional Distributed Source Field Problems

EDWARD O. GILBERT
NONMEMBER AIEE

ELMER G. GILBERT
NONMEMBER AIEE

POISSONIAN or distributed source fields occur often in present-day problems dealing with such subjects as torsional stress analysis, heat transfer, viscous fluid flow, eddy currents, and electric and magnetic fields. Since the mathematical solution of Poissonian fields has only been possible in several special cases, physical analogues have been developed for 2-dimensional fields. A highly successful physical analogue is the sandbed fluid mapper.[1-3] This paper describes an electric analogue mapper which was originated and developed by the authors. The analogue can give directly usable 2-dimensional solutions of high accuracy.

The Analogue Field Map

The mapper utilizes a resistive plotting surface and a coupling capacitor. The coupling capacitor is formed by the resistive surface and a parallel coupling electrode. A displacement current acting through this coupling capacitor sends a distributed current into the resistive surface. As shown in the Appendix, this distributed current flow represents Poisson's equation. Boundary conditions are introduced by suitable conducting terminations of the resistive surface. The coupling electrode shape and area correspond to those of the distributed source. The displacement current density is readily controlled by the spacing of the coupling electrode and the resistive surface. Most of this paper will discuss mappers of constant displacement current density.

In practice, Teledeltos recording paper (manufactured by the Western Union Telegraph Company) is used as a resistive surface. A dielectric material separates it from the coupling electrode, a conducting sheet of aluminum foil or silver paint. Silver paint is used for the conducting boundary of the Teledeltos paper. A circular map made of these materials is illustrated in Fig. 1. A uniformly distributed source region is represented since the dielectric is uniform in thickness. This map could, for example, represent the magnetic field within a circular conductor of infinite length. In operation the terminals of the mapper, that is, the conducting boundary line of the resistive paper and the conducting sheet, are connected to a source of alternating voltage. The resistance of the paper is small in comparison with the reactance of the coupling capacitor. For this reason the voltage drop across the resistive paper is a small percentage of the applied voltage and therefore a substantially uniform source of current density appears in the resistive paper over the conducting sheet. This uniform current flow represents the problem of the uniformly distributed source field. A null-type measuring circuit is used in conjunction with a probe to plot the equipotential lines on the resistive paper. These lines correspond to one set of the orthogonal system that is the solution of the desired field.

For the capacitively coupled field mapper to be a workable analogue, the current density appearing in the resistive paper must be inversely proportional to dielectric thickness; this requires that the voltage across the dielectric be the same at every point. However, the mapper works on the principle of a measureable voltage drop across the map surface. The error caused by this imperfect analogy was mathematically investigated for the circular map and found to be entirely negligible for any practical mapper.

Equivalent Circuit

A simplified circuit representation of the mapper would be useful in further work. An equivalent circuit is shown in Fig. 2. It has proved to be highly accurate for all maps thus far measured. The impedance of the coupling capacitor C is generally 100 to 500 times as large as R_S, which represents map surface resistance. The resistance R_L is apparent in measurement and is believed to denote loss in the coupling capacitor. Typical values of R_S range from 400 to 4,000 ohms. The point of maximum voltage on the resistive surface is the kernel. It is point K on the diagrams. The maximum surface voltage E_k will be referred to as the kernel voltage.

Measurement Methods

The measurement of equipotential lines presents difficult problems which may best be understood by examining conditions under which a typical mapper might operate. The resistive paper is held near ground potential while the coupling electrode is excited to a potential of about 100 volts. With average conditions, less than 1 volt will appear across the map surface. In weak field regions the voltage gradient is extremely low and precise measurements must be made if points are to be located accurately. Minor changes in C or R_S can produce large changes in the measured position of equipotential lines. For this reason the stability of mapper components is extremely important. Capacitive drift (change in coupling capacity with time) and resistive drift (change in paper resistance with time) do exist and increase the difficulty of accurate measurement. These effects will be discussed more fully later.

All of the measuring circuits tried were of the null type. A vacuum-tube voltmeter would be ineffective because of the low voltages present and accuracy required. A null-type detector has proved convenient because it can give an audible signal not demanding that the operator be continually shifting his vision from the plotting surface to a visual indicator. The low-voltage levels present indicate that steps should be taken to minimize stray couplings. This is shown to be true in actual practice, as without proper techniques deep nulls are impossible. The circuit diagrams will show where shielding is employed. A guard electrode placed near the high-potential coupling electrode is invaluable in reducing stray flux. Such a guard electrode is shown in Fig. 1. A sensitive null indicator is required in all cases. Head-

---

Paper 53-193, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Southern District Meeting, Louisville, Ky., April 22–24, 1953. Manuscript submitted October 21, 1952; made available for printing March 6, 1953.

Edward O. Gilbert and Elmer G. Gilbert are both at the University of Michigan, Ann Arbor, Mich.

The authors are especially indebted to Prof. A. D. Moore for his help and encouragement. Thanks are also due to Prof. M. B. Stout and the many others at the University of Michigan who have helped make this paper possible.

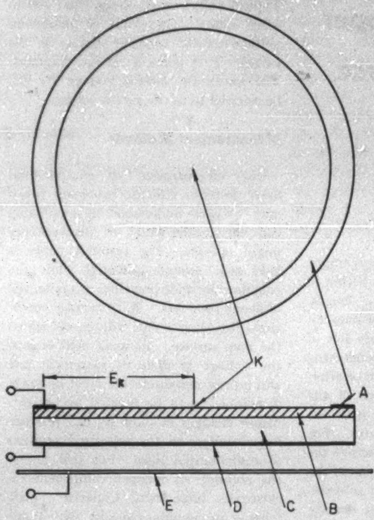


Fig. 1 (left). A circular map with guard electrode

A — conducting boundary
B—resistive paper
C—dielectric
D — conducting sheet
E—guard electrode
K—kernel

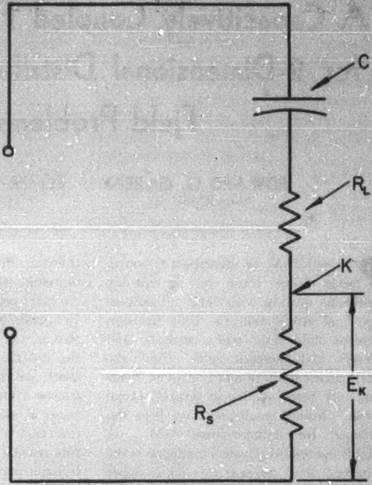


Fig. 2 (right). Equivalent map circuit

C—coupling capacitance
R_L—loss resistance
R_S—surface resistance
K—kernel

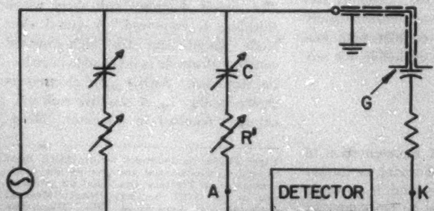
phones in conjunction with a battery-operated bridge amplifier make a good combination. A selective filter in the detector amplifier circuit is helpful in reducing the effects of hum pickup from nearby a-c lines. A sharp needle is the best probing tool since a dull point makes positive surface contact difficult. These points are applicable to all of the circuits used. Map requirements have led to the development of three basic types of measuring circuits.

The first circuit is shown in Fig. 3. The variable circuit elements *R′* and *C* are adjusted so point *A* has the same potential as the kernel. Any fraction of the kernel voltage may then be measured by setting the slider of *R* at the desired percentage of full kernel voltage. A Wagner ground circuit is used to eliminate the effects of stray detector capacity to ground. This circuit produces good nulls and satisfactory accuracy but requires bothersome adjustment of the Wagner ground circuit for each different potential measured. Capacitive and resistive drift also cause troublesome shifting in the measured position of equipotential lines. For these reasons the next measuring method is preferred.

The transformer circuit, shown in Fig. 4, gives excellent results. Not only does it use less equipment than the previous circuit but it is able to counteract the effects of capacitive drift in the map.

Fig. 3 (below). Measuring circuit with Wagner ground

C_S—stray detector capacitance to ground
G—guard electrode
P—shielded probe



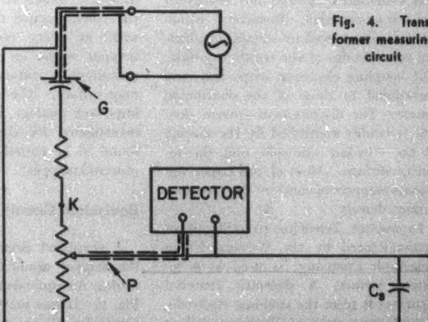
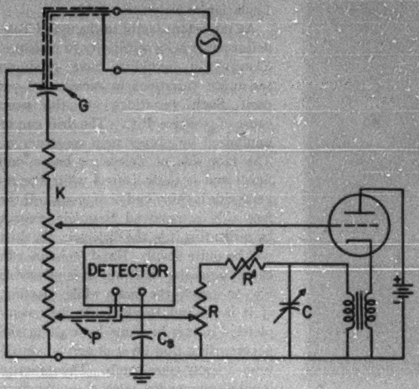


Fig. 4. Transformer measuring circuit



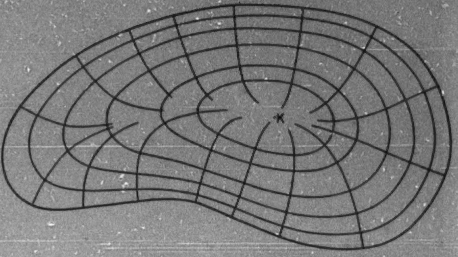


Fig. 5 (left). Vacuum-tube measuring circuit

Fig. 7 (above). The solution of a uniformly distributed source field by two methods. The closed curves are equipotentials obtained from the capacitively coupled mapper. The set of flow lines was obtained from a sandbed fluid mapper. The excellent orthogonality of the sets shows close agreement between the two solutions

The transformer gives a 180-degree phase shift so that point A may be adjusted to the same potential as the kernel. Capacitor C serves to correct phase shift in the transformer and its associated circuit. Since the equivalent capacitor of the map has by far the highest impedance of the series circuit, it is the main factor in controlling the total current through the map and the transformer primary. Thus, the voltage across R may be conveniently adjusted by changing R'. If the reactance of the map capacitor should increase slightly due to drift, then the current through the circuit would decrease slightly. This in turn would cause a slight drop in the kernel voltage and in the voltage across R. Since the potential of the kernel and point A rise and fall together, capacitive drift has no effect on circuit balance. This method does not compensate for the effects of resistive drift which may be minimized by proper handling of the resistive paper. The circuit is easy to adjust as there is only one control to change in measuring different potential lines. The stray capacity C_S can cause trouble if R is too large. Operation has been satisfactory if R is kept below about 500 ohms. This can easily be done by governing the turns-ratio of the transformer. In all maps measured a turns-ratio of n_1/n_2 equal to four has been satisfactory. A bridge-type transformer provides excellent isolation. It should be noted that the guard electrode is not returned to ground but is shunted across the oscillator. This has been found necessary to make certain that the transformer carries current governed only by the map capacitor; otherwise, the circuit would not completely counteract capacitive drift.

A vacuum-tube circuit is shown in Fig. 5. The measuring resistance R takes the form of a load in a cathode-follower amplifier. The step-up transformer is necessary since the voltage gain of the stage is less than unity, and since the cathode-follower input voltage is less than the kernel voltage. The grid terminal of the tube is kept in contact with the map somewhere near the kernel. Map loading is negligible because the input impedance of the tube is exceedingly high. Since the voltage across the measuring resistance varies as the voltage across the map, the circuit responds to neither capacitive nor resistive drift. This circuit is effective in operation only if the tube potentials are carefully stabilized. The measuring method has all the advantages of the transformer circuit, but does require more equipment.

The transformer circuit has been found very satisfactory and has been used in the majority of measurements. With proper control of resistive drift it is accurate, simple, and easy to use. A typical setup is pictured in Fig. 6. There are, no doubt, many other methods which will give excellent results. The examples given show what can be done with suitable circuitry. The choice of method will depend upon the user's requirements.

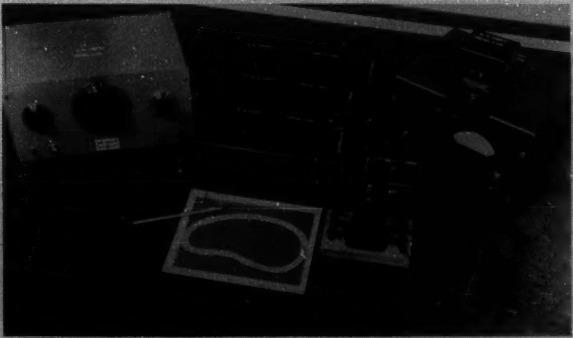


Fig. 6. A typical experimental setup. The transformer measuring circuit is used

Causes of Error

Errors in line positions are due to five causes: inaccuracy of measuring equipment, voltage drop across the map surface, nonuniformity of the dielectric, nonuniformity of the resistive paper, and boundary fringe flux. As mentioned before, errors due to surface voltage drop are extremely small in a practical mapper. This type of error is roughly proportional

Fig. 8. Actual mapper representing the magnetic flux within a rectangular duct of infinite permeability and length. The duct holds two rectangular, symmetrically placed conductors with equal, uniform current densities in the same direction. Shown is the left half of the duct with the centerline at the right. One of the conductors can be seen in the lower left corner. Shown are flux lines and the kernel K

to distributed source area, the resistance per square of the paper, and the admittance per unit area of the capacitor. For a circular map of moderate area (100 square inches), thin paper dielectric (several mills), and usual frequency of operation (1,000 cycles per second), the maximum error in line position from this source is less than 0.01 per cent of the map diameter. Since the upward current density is dependent upon the dielectric thickness and the dielectric constant, these factors must be carefully controlled if accuracy is to be achieved. A homogeneous dielectric of high dimensional accuracy is necessary. Care must also be taken to insure the uniformity of the adhesive layers which are used to assemble the map. Another source of error is resistive paper nonuniformity. The nonuniformities of Teledeltos paper will be discussed in some detail later. At the edge of the conducting sheet the flux to the conducting paper will not change abruptly because of fringe flux effects. This fringing can cause a slight error at the distributed source boundary. It is believed that this error is extremely small in a practical map where the dielectric thickness is small compared to map area. When boundaries of the resistive paper and the coupling electrode coincide, the fringe flux may be eliminated by extending the coupling electrode. This has been done in Fig. 1.

Conducting Paper

Teledeltos paper has proved to be a good resistive surface medium. It is readily prepared and maps of great complexity can be built quickly at negligible cost. Conducting boundaries are simply applied using silver paint. The paper is available in two types, high and low resistance. The low-resistance paper, type *L-39*, has a resistance of approximately 2,000 ohms per square and is used because it gives less surface voltage drop.

Teledeltos paper is not perfectly uniform in its resistive properties. The major variation in resistance seems to depend on direction of current flow. A maximum difference in resistance occurs when current flow is first parallel and then perpendicular to the direction in which the paper was rolled. Variations as high as 16 per cent and as low as 2 per cent have been measured in different samples. If selection of the paper were possible, this error could be eliminated or greatly minimized. The error in line position caused by this variation depends upon map geometry. In most cases the error is very small. Other relatively minor nonuniformities are believed to exist. It is hoped that a resistive paper of high uniformity will be made available for mapping purposes. Such a paper would have great value for mapping either Laplacian or Poissonian fields.

An important factor in the use of Teledeltos paper is its sensitivity to moisture. Changes in humidity have produced resistance variations in excess of 1 per cent. Such variations are the main cause of resistive drift. The drift can be minimized by proper map construction. The face side of Teledeltos has a dull finish and is quite porous while the reverse side is shiny and nonporous. If the face side is removed from air contact by cementing it to the dielectric, the drift will be quite small. Hand moisture can cause noticeable drift and it is advisable not to touch the map while plotting. If it is desired to eliminate drift completely, it can be done by coating the paper with a thin layer of acrylic plastic from a spray-type bomb. This practice is not recommended unless necessary since it makes probing quite difficult. In most cases the first precaution is entirely adequate.

Dielectric Material

There is a problem in finding dielectric materials that are homogeneous and dimensionally accurate. The problem is further complicated if an effort is made to eliminate capacitive drift. The form of mapper construction makes it extremely difficult to construct a stable coupling capacitor. It is better to eliminate the effects of capacitive drift by suitable measuring circuitry than to eliminate the capacitive drift itself. If high accuracy is not needed, common dielectrics such as paper and plastics may be used. A good grade of window glass is an excellent material since it is dimensionally accurate (within a few per cent), resistant to attack of adhesives, and easy to obtain. The final choice of a dielectric will depend on how severe the user's accuracy requirements are.

Map Construction

To illustrate possible map-making techniques, the construction of a uniformly distributed source mapper will be described. The dielectric, in this case a sheet of glass, is sprayed with acrylic plastic until it is just wet. The Teledeltos paper is then applied and rolled smooth. Considerable pressure on the roller is necessary to remove excess plastic and to insure a uniform plastic film. The uniformity of the film may be checked easily by viewing the paper through the glass. Since the plastic sets fairly rapidly, some speed is required in executing these steps. This method of assembly is fast and produces results far superior to those using the more common cements, glues, or pastes. The coupling electrode may be of either silver paint or aluminum foil. The foil is quickly applied, using the same technique as is used with the Teledeltos paper. If small bubbles appear they can be removed easily, by piercing them with a pin and then rolling them flat. Correct foil uniformity will result in a mirror-like surface. The foil should then be trimmed so that it covers only the area where distributed flow takes place.

The silver paint is useful in the patching of torn foil. Silver paint may replace the foil but its application is more time-consuming and costly. The map is then ready for its associated wires, conducting boundaries, and guard electrode.

Conducting boundaries are easily affixed to the Teledeltos paper with silver paint. The silver paint is a commercially available product used in printed circuit work. Where high accuracy is desired, the paint may be applied with a ruling pen. A boundary width of 3/8 inch has been found adequate. Tape is used to hold the wires in contact with the map boundaries. Electric connection is then made by daubing the wires to the boundaries with silver paint. To assure high conductivity such connections are made every few inches. The wires may be held more securely if they are kinked or looped. Connections to the coupling electrode are made in the same manner. To protect the map and its connecting wires, it is helpful to mount it on stiff cardboard. A foil guard electrode is cut slightly larger than the coupling electrode and is affixed to the top of the mounting cardboard with cement. A lead is then attached to the guard electrode and a hole punched through the cardboard and guard electrode for the capacitor lead. A sheet of paper is used to separate the guard electrode and the bottom of the coupling electrode; the components are then assembled and taped together. When the shield wire is attached and secured to the mounting cardboard, the mapper is ready for operation.

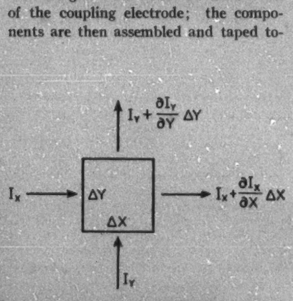

$I_Y + \frac{\partial I_Y}{\partial Y}\Delta Y$
I_X
ΔY
ΔX
$I_X + \frac{\partial I_X}{\partial X}\Delta X$
I_Y

Fig. 9. An area in the xy plane used in the development of Poisson's equation

Results

The accuracy of the capacitively coupled mapper is good. A circular map as in Fig. 1 was constructed with a glass plate dielectric with a diameter of 20.0 centimeters. The maximum error in equipotential line position was less than 0.1 centimeter. Average error was much less. Complex fields have been checked with sandbed fluid mappers. Such a check is shown in Fig. 7. The fluid map is by Prof. A. D. Moore. Excellent orthogonality of equipotential lines and representative fluid flow lines indicates close agreement. The fluid flow lines also seem to originate from the kernel which was found by the capacitively coupled mapper.

Fig. 8 shows an actual mapper as used to solve a magnetic field problem. Note how the symmetry of the problem enables the mapper to be simplified. The current enters the resistive paper in the distributed source area at the lower left. It leaves from the conducting boundaries at the top, left, and bottom. Here the equipotentials represent magnetic flux lines. No refraction of flux lines occurs at the distributed source boundary. This is as it should be when the permeabilities are the same on either side of the boundary.

Conclusion

At present the accuracy of the mapper is limited by resistive paper and dielectric nonuniformity. Even so, results obtainable are excellent. The maximum error in line position should not exceed 1 per cent of the maximum map dimension and average error should be much less. The time involved in making and using a mapper such as the one described is in the order of 3 to 4 hours. The value of any potential line can be determined easily. In most cases the equipotential lines are the ones desired in the final solution.

The mapper should easily be applicable to regions of nonuniform displacement current density. Though such a mapper has not yet been constructed, little trouble is anticipated. The method of capacitive coupling might also be applied to the electrolytic tank, by-passing the problem of nonuniform resistive paper. The electrolytic tank might also make possible the solution of 3-dimensional fields with axial symmetry.

Appendix

If the capacitively coupled field mapper is to solve 2-dimensional fields of the distributed source type, it must satisfy Poisson's equation. In the following development the voltage drop across the resistive paper is considered very small in comparison with the applied voltage. Therefore the displacement current density appearing in the resistive paper is inversely proportional to the dielectric thickness. Let J equal the displacement current density at any point in the distributed source area. Let K equal the resistance per square of the resistive paper. Consider the sum of the currents leaving a small area in the xy plane as shown in Fig. 9.

$$\frac{\partial I_x}{\partial x}\Delta x+\frac{\partial I_y}{\partial y}\Delta y+J\Delta x\Delta y=0 \tag{1}$$

The voltage drops across the area are

$$\Delta V_x=\frac{\partial V}{\partial x}\Delta x=I_xK\frac{\Delta x}{\Delta y}$$

$$I_x=\frac{1}{K}\Delta y\frac{\partial V}{\partial x}$$

$$\frac{\partial I_x}{\partial x}=\frac{1}{K}\Delta y\frac{\partial^2 V}{\partial x^2} \tag{2}$$

$$\Delta V_y=\frac{\partial V}{\partial y}\Delta y=I_yK\frac{\Delta y}{\Delta x}$$

$$I_y=\frac{1}{K}\Delta x\frac{\partial V}{\partial y}$$

$$\frac{\partial I_y}{\partial y}=\frac{1}{K}\Delta x\frac{\partial^2 V}{\partial y^2} \tag{3}$$

Substitute equations 2 and 3 in 1 and the result is

$$\frac{\partial^2 V}{\partial x^2}+\frac{\partial^2 V}{\partial y^2}=-KJ$$

which is Poisson's equation for two dimensions.

References

1. FIELDS FROM FLUID FLOW MAPPERS, A. D. Moore. *Journal of Applied Physics*, New York, N. Y., vol. 20, Aug. 1949, pp. 790–804.

2. SOAP FILM AND SANDBED MAPPER TECHNIQUES, A. D. Moore. *Journal of Applied Mechanics*, New York, N. Y., vol. 17, Sept. 1950, pp. 291–98.

3. THE FURTHER DEVELOPMENT OF FLUID MAPPERS, A. D. Moore. *AIEE Transactions*, vol. 69, part II, 1950, pp. 1615–24.

Discussion

Walter E. Rogers (University of Washington, Seattle, Wash.): This is an important paper, and the Gilbert twins are to be commended for their contribution to a field in which there is such general interest. I look forward to the extension of their development to nonuniform distributed sources, something which has not been achieved with the sandbed fluid mapper or the membrane analogue.

I would like to point out that while the electrical measurements discussed here are rather elaborate and time-consuming, they are certainly not more so than are the presently used methods[1] of measuring the displacement of a soap film or rubber membrane; nor is the electrolytic tank without measurement difficulties. The capacitively coupled field mapper is superior to the sandbed in the matter of fine detail and in the exact location of the kernels. This is an opinion which would be substantiated, I think, if the authors and Professor Moore would submit a photograph of the original sandbed map of Fig. 7 of the paper with their answer to this discussion.

Analogues are important academically because they are an invaluable aid in the challenging problem of teaching students to visualize potential fields. We plan to include the Gilberts' development along with Professor Moore's fluid mappers and some membrane analogues which are presently used in our laboratories at the University of Washington. To this end, would the authors please state how the loss resistance R_L of their equivalent circuit becomes apparent in the measurements? Is it that points A and K in Fig. 3 of the paper are not at the same potential unless R' is included?

There is one other detail in the paper which needs clarifying. Are the equipotentials, which surround the kernel in Fig. 8, taken at equal increments of potential with respect to that of the kernel? In the region of the kernel, the gradient seems small in comparison with Fig. 7. I observe, however, that a vertical line through the kernel and a horizontal line to the left of the kernel are both stream lines. This indicates that the kernel is correctly located.

Reference

1. See reference 2 of the paper.

J. H. Fooks (Westinghouse Electric Corporation, Sharon, Pa.): The authors are to be commended on their excellent contribution to experimental mapping methods. The thoroughness of their investigation with regard to errors involved and, particularly, their solution of the capacitive drift problem via the transformer measuring circuit should clear the way for the immediate application of this method to industrial field problems.

The possibility of simulating nonuniform displacement current densities by varying dielectric thickness was of particular interest. This is readily accomplished in a fluid mapper by using a sandbed of varying depth. Offhand, I cannot see how this could be done easily in a capacitively coupled mapper using a reliably uniform dielectric material such as glass. Considerable care would have to be given to the selection of a more workable material. Perhaps the authors would care to say how this might be done.

The time involved in constructing and operating a capacitively coupled and a fluid mapper is, perhaps, much the same if the fluid mapper is carried through to a finished photograph of the field.

Almost all of our application of experimental mappers has been confined to 3-dimensional fields of axial symmetry. Fluid mapping is readily adaptable to this case by varying the flow spacing, whereas the capacitively coupled mapper must await the advent of a suitable resistance paper. It is not inconceivable that such a paper might be forthcoming. The capacitively coupled 2-dimensional mapper might be used to approximate the solution of a 3-dimensional field or to show field trends. This has frequently been done mathematically to simplify methods of solution.

In the past, the tools of experimental mapping methods and field mapping have not been used to any great extent in field instruction or problem solving. This has been unfortunate since the visual analysis of fields and field trends provides the quickest and most lucid solution of field phenomena with reasonable accuracy.

Within the past several years experimental mapping has received a tremendous stimulus through the excellent work of Professor Moore in the invention and development of fluid mapper methods. It is gratifying to see the interest in this development and the trend in new text material to include sections on practical mapping methods.

It is to be hoped that, with the availability of fluid mappers and similar experimental methods such as described in this paper, the visual simulation of field phenomena will receive its due recognition.

A. D. Moore (University of Michigan, Ann Arbor, Mich.): Learning about sandbed fluid mapper methods for simulating the distributed source, the authors soon turned to the invention and development of the new analogue they have so ably reported on in their paper.

The equipotential lines obtained by their method are orthogonal to the flow lines obtained over the sandbed, in a given case. They have applied their method to several sandbed cases already available from the writer's work, and came out with a high degree of correspondence. This, by the way, constitutes a check on both methods.

Professor Rogers indicates that simulation of nonuniform distributed sources has not been achieved with sandbed fluid mappers. Rather, it has been achieved but not yet reported. The writer has built a slab with a rectangular sandbed, in which a sloping screen makes the sandbed depth vary by a ratio of 3 to 1 from one end to the other. The flow density appearing on top of the bed then is nonuniform, varying approximately inversely as the sandbed depth. In addition to sloping the screen, there are at least two other methods for achieving nonuniform distributions with the sandbed.

The Gilbert contribution is a welcome addition to our knowledge dealing with fields. Field problems constantly arising are of such range and complexity that we need all of the analogues we can get.

J. F. Calvert (Northwestern University, Evanston, Ill.): The authors are to be congratulated on the development of what I believe to be an entirely new approach to the solution of 2-dimensional fields with distributed sources. It would appear that they have devised a method which can be readily used by any engineer and scientist with normal skill in electric circuit and measurement techniques, and it would seem safe to predict that this method will be generally accepted. I hope the authors will (1) show several more examples of the results obtained by this technique; (2) discuss the possible extension of the equipment to 2-dimensional problems where the distributed sources are not of uniform strength; and (3) further discuss the possible applications to 3-dimensional fields, including the probable difficulties and limitations as they may see them at this time.

W. R. Simmons (Argonne National Laboratory, Chicago, Ill.): The analytical solution of problems involving Poissonian or distributed source fields has always been a long and tedious task except for those few cases where known solutions exist. The method described in this paper will expedite the solution of many of these 2-dimensional problems. This method should be of interest to the engineer since most practical engineering problems involving Poissonian fields do not have known solutions. Further, it should be noted that the materials required to set up a problem by this method are inexpensive and answers with sufficient accuracy for engineering work can be rapidly obtained.

The method described in this paper may be particularly valuable in solving 2-dimensional heat transfer problems which involve Poissonian fields. This type of problem constantly confronts the nuclear heat transfer analyst. By suitable construction of the analogue, heat transfer problems involving any or all of the following conditions can be solved: (1) Poisson fields; (2) Laplace fields; (3) materials of various thermal conductivities; and (4) convective heat transfer from a surface to a fluid.

Poisson fields can be built up in a number of ways. However, the method devised by the authors is one of the more ingenious methods. The Laplace fields are obvious since only conduction is involved.

Materials of different thermal conductivities (or resistivities) can be simulated by constructing the analogue so that more than one path of the same shape is available for the electric current. These parallel paths can be electrically connected by carefully applying silver paint in a discontinuous manner around the boundaries.

Convective heat transfer from a surface to a fluid can be easily included in an analogue constructed from Teledeltos recording paper. This is one of the principal advantages of an analogue of this type. This is accomplished by adding to the surface where convective heat transfer takes place an additional length of Teledeltos recording paper. This additional length should be equal to the ratio of thermal conductivity to the heat transfer coefficient (k/h) in order to properly represent the film resistance. This additional length (k/h) of conducting paper must be cut in strips perpendicular to the surface from which the convective heat transfer occurs. This insures an electric current flow which will be orthogonal to this surface. In this manner problems involving 2-dimensional Poissonian fields and convective heat transfer from nonisothermal surfaces can be readily solved.

The heat transfer section of the Argonne National Laboratory has developed a similar type of analogue. However, we approached the problem somewhat differently. We are primarily interested in actual temperature differences associated with certain heat flows through Poisson and/or Laplace fields with convective heat transfer from a nonisothermal surface. Therefore, we have developed a d-c system which gives us directly values which are proportional to temperature differences and does not involve evaluating a conformal map.

Edward O. Gilbert and **Elmer G. Gilbert**: The authors appreciate the interest shown in the capacitively coupled mapper. It would be well to clarify the points mentioned by Professor Rogers. The potential lines in Fig. 8 of the paper have values of 0.8, 0.6, 0.4, and 0.2 E_k with the singular point at 0.246 E_k. Potential lines in Fig. 7 are at 0.9, 0.8, 0.6, 0.4, and 0.2 E_k. The presence of R_L is directly known from the settings of R and R' in Fig. 3. It is apparent that R_L is not detected by the other circuits. Note that K and E_k are not known initially, but must be determined by a series of rapidly convergent trials.

The resistive paper modifications discussed by Mr. Simmons should prove very useful. Values of resistance per square had by paralleling resistance sheets are limited to discrete steps. An array of small, uniformly spaced conducting dots would provide another approach to the problem. Such an array might be applied to the resistive paper using a stencil and silver paint. It is hoped that more versatile resistive materials will be developed by someone in the future. The electrolytic tank could handle wide changes in resistivity with some increase in physical complexity. Capacitive coupling to a tilted tank would facilitate the solution of 3-dimensional problems with axial symmetry. Here, the dielectric material would form the bottom of the electrolytic tank. The additional problems encountered should not be too great as electrolytic tank methods are well advanced.

The nonuniform distributed source can be generated in several ways. A multilayer dielectric could be formed in contour fashion to provide a step approximation to a desired thickness function. Dielectrics such as celluloid, vinyl plastic, and paper have proved reasonably successful in constructing uniform source mappers; their extension to multilayer dielectrics should be convenient and inexpensive. Displacement current errors of less than 10 per cent are foreseeable. Even better accuracy would be possible with machined dielectrics. A subdivided coupling electrode and a uniform dielectric would allow a more versatile approach. Each segment of the coupling electrode could be assigned a desired current. One mapper could then solve problems with many different distributed source configurations.

The authors hope that this brief discussion will answer the questions of Professor Calvert and Mr. Fooks. Development of the modifications mentioned should greatly increase the usefulness of the capacitively coupled field mapper.

Reliability in Industrial Electronic Equipment

ELLSWORTH D. COOK
ASSOCIATE MEMBER AIEE

THE EXPECTATION of the purchaser of industrial equipment is no different from that of the purchaser of any other form of equipment or, in fact, from our own when we purchase things for use or pleasure. Reliability is expected. Such reliability, however, includes more than long life, important as that may be; it also means long periods of freedom from readjustment and similar annoyances. The following observations relative to factors which enter into such reliability were obtained from experience in the field of industrial electronics, especially as found in the heavy industries. A word of caution is in order at this point: The impression that electronic equipment is hopelessly involved with repeated failures is not a true one. What is reported as in need of correction is actually based upon the known facts that performance to date has been such as to commit industry to the use of electronic means, and that the desired improvements are possible and worthwhile.

It is unfortunate that no magic formulae exist by which equipment can be produced with the desired reliability. The means of accomplishing the objective are known to all of us, but have something in common with certain medicines in that we would prefer not to have to use them. Like medicines, since the ideal ones are not yet available, we use those remedies we know and have.

Since sufficient improvement in component reliability has not yet occurred which automatically would permit a major improvement in equipment reliability, engineers must use such expedients as their combined experience shows to be helpful. These are the simple, commonsense, but totally unromantic methods of established good engineering practice in design, installation, maintenance, and service. What must concern engineers for the present, therefore, is continued emphasis on these practices and further extension of their use. While all of these functions are important, at the present time most emphasis must be placed on good design; first, of the component and, second, of the equipment. However, because the methods cited for the equipment as a whole have a limited effectiveness, emphasis must be given to the encouragement by word and other support to the component manufacturers in their efforts to improve the reliability of their products.[1]

With so many electronic engineers involved in military developments, it sometimes comes as a surprise that the field of industrial electronics is experiencing a substantial expansion. With this expansion, the need for greater emphasis on reliability is increased. There is a need for wider understanding of what makes one piece of equipment reliable and another unreliable in each field, and there is a need for better knowledge of how to keep good equipment in reliable operation. While the technical subject matter is not materially different in the various fields of applied electronics, the emphasis placed on the various elements do differ to a marked degree. Thus, in the broadcast field, it is essential that the program stay on the air; but in the power field, where production, and even life itself, is sometimes jeopardized by failure of equipment, the need for guaranteed performance reaches a new order of magnitude. Perhaps the fact that in America the clocks of so much of the nation now depend upon the power system will illustrate this expected freedom from failure. In heavy industry, experience has shown that failures are most frequently due to faults of a mechanical nature. Thus, mechanical design is likely to be every bit as important as the purely electronic aspects of the device, and designs which

---

Paper 53-208, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 6, 1953; made available for printing April 8, 1953.

Ellsworth D. Cook is with the General Electric Company, Schenectady, N. Y.

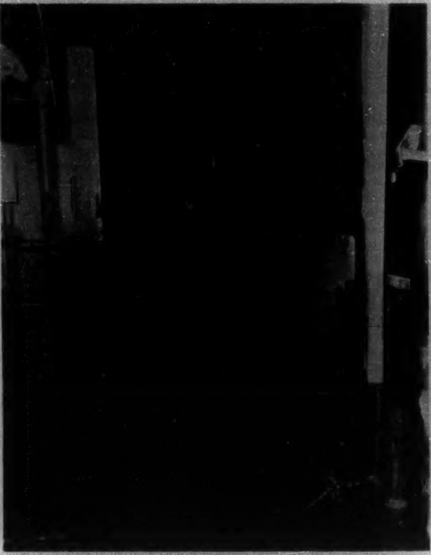


Fig. 1. X-ray thickness gauge at delivery end of hot-strip steel mill at Irvin Works of the United States Steel Company

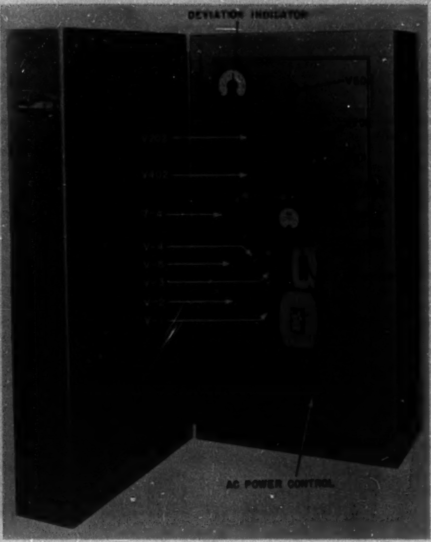


Fig. 2. Electronic cabinet—width gauge for hot-strip steel mill

would be mechanically acceptable in the radio field cannot be tolerated in heavy industry.

Design Aspects of Reliability

It is not of so much importance that a unique and efficient electronic circuit be provided as it is that the equipment be able to stand the every-day treatment which is normal in that industry. The use of such industrial equipment dictates many aspects of the design insofar as the desired reliability is concerned. Choice of materials which may be used and knowledge of methods of fabricating often assume new importance. Equipment which must be located beneath the pass line in a hot-strip steel mill will be subjected to heat and a bath of water, steam, and scale, and in many installations the water will, without doubt, be contaminated with acids. From Fig. 1, the normal habitat of such equipment may be seen. Unquestionably a stainless-steel tank would serve admirably as the cabinet, but if this tank must also hold light insulating oils, it is by no means certain that stainless steel should be used. The question of how to weld such a structure to retain the oil would now assume a far greater importance than whether or not to use degeneration in some amplifier within such a cabinet.

Ordinarily, enclosing cabinets for electronic equipment would not be considered in a discussion of reliability. However, if they are not reasonably dust tight, or are not built so that parts are readily accessible for service and inspection, these functions are likely to be slighted to the detriment of the performance. Fig. 2 shows a recent improvement that offers much in that regard. The swinging inner panel, the large flat surface to apply dust and moisture sealing gaskets, the strong handle which closes the cabinet under great pressure against these gaskets, and the minimum of cracks and openings have been found to be definite improvements over cabinets used previously in industrial applications.

In examples such as the above, another design problem can arise. Totally enclosed cabinets can accumulate considerable moisture due to repeated alternate heating and cooling. Cases have been found where the connection cables breathed, permitting moisture to enter. This entrapped moisture can lead to equipment failure in many ways, not all of which are immediately obvious, but one of the more well-known ones is due to creation of acids. In this connection, moisture-absorbing drying agents are not always an adequate solution, and, in certain industrial equipment it has been found profitable in some applications to keep a cabinet interior under pressure with moisture-free inert gases, as a means of increasing the reliability of performance.

The design of mechanical parts, methods of fastening, and support of components are of importance in all products, but these considerations must receive different treatment in different fields. The mechanical stops used on certain switches and similar parts may be entirely acceptable in certain fields, but in industrial use, these can be the reason why that part is not acceptable, and this fact must be considered by the designer in obtaining system reliability. Fig. 3 shows an example of an industrial power switch which has provided good service in electronic applications in industry. Its sturdy handle, stops, and other parts are especially noticeable. In one field, it may be just as important to successful operation to use good insulation to support high-voltage components as in another, but where heavy industry is concerned a different approach is generally needed. The configuration of insulating material may be of no consequence in certain devices, while in others operating at high voltage, it may be a

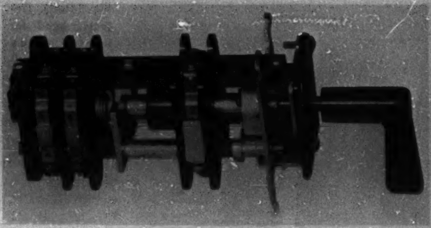


Fig. 3 (above). Heavy-duty rotary switch for industrial use

Fig. 4 (right). First installation of noncontacting width gauge on hot-strip steel mill at Irvin Works

cause for ionization currents and frequent breakdown.

Several examples will show the nature of these differences in approach. The crowded component boards of certain miniaturized electronic equipment are a hazard in industrial electronic devices. Larger creepage distances are desirable. Where a tapped hole may be used in a piece of fabricated insulation of the resin-bonded variety and a component may be fastened to it by means of a machine screw in equipment which will be handled with care, equipment intended for applications where the vibration of heavy industry is involved requires either a threaded metal insert, so installed as to avoid splitting of the insulation under normal operating conditions, or a clearance hole through the insulator with metal clamp pieces on either side must be provided. Shock-mounted components may be necessary to avoid microphonics in almost any field of applied electronics, but the materials which serve for communications work cannot always be used, without incurring a service problem in industrial electronics. Materials like rubbers do not maintain stable positions of rest and when subjected to heat, strain, and certain atmospheres, sometimes lose their desirable characteristics. In such instances, spring filter systems and mechanical-type shock absorbers are often necessary. In other cases, in which an electronic engineer might ordinarily employ shock mountings, equipment must be operated so that certain components will maintain fixed positions relative to other devices. Here other means to eliminate the effect of vibration must be found. For example, it was desired to use an X-ray tube, originally designed for diffraction work, in an electronic device for cold-strip steel mills. The massive parts, mounted on a glass structure, would not tolerate the increased vibration encountered and, because the X-ray source had to maintain a fixed relation to the detector unit, shock mounting was impossible. In such cases, either a different or a redesigned tube would have to be employed. Fortunately the latter alternative was possible.

The use of some methods of mounting parts, such as those employed with certain transformers, where a few holes are provided in a thin metal base, and which might be acceptable in some fields of electronics, would be totally unsuited in the heavy industry field. The transformer case would have to be redesigned for the more severe handling it would encounter. Heavy tubular-paper condensers which are sometimes mounted on their connection wires in a radio chassis, or secured by means of tying or a compression clamp in other applications in this field, should be enclosed in a container which would provide greater mounting security.

Bolts which might be adequate to secure a part in ordinary use frequently have to be increased in size all out of proportion to the strength requirements, because of the frequency with which they may have to be removed, or because the service personnel is not accustomed to treating bolts with the respect of an instrument maker. Not infrequently a designer has found that if a large number of small bolts is required to secure a part, the threads of either the bolt or its mating part are stripped in service and, because of the difficulty of repairing the damage, that bolt is not replaced, with the result of ultimate failure of the component involved. Such cases arise where a cover and gasket are used to close an oil-insulated component or a pressure chamber. Regardless of the fact that certain methods of caring for the circumstance could have been found which would have prevented the failure, ultimately the orders will go out for that manufacturer's equipment, if it can be found, which was designed to avoid that circumstance. For similar reasons, it is found, not infrequently, that otherwise useful equipment will be unused if too much trouble of this sort is experienced.

Another problem frequently encountered in the design of reliable industrial electronic equipment, which is not common to most other electronic designs, is that of lubrication. Industrial electronic equipment frequently requires motors of fairly respectable size to move its parts, or has accurate mechanical parts such as gear systems, levers, ways, slides, or guides which demand metals having certain wearing properties. Certain of these may require the use of steel which should not be plated. Such items require lubrication for one reason or another. An example may be in order: It was necessary in a recent industrial device to move two assemblies weighing about 40 pounds each with some appreciable accuracy in regard to position and path of travel. The installation of this equipment is shown in Fig. 4. Therefore, the supporting member upon which this motion had to take place had to be strong, rigid, and resistant to wear. Steel was the final choice of material, and the surface could not be plated to prevent rust and corrosion. This structure is shown in Fig. 5. Therefore, it was necessary to provide lubrication, not so much because it was desired to reduce friction, although this was important, but mainly to prevent rust and corrosion. Since the device was exposed alternately to long periods of high temperature, followed by a shorter period of low temperature, sometimes in the freezing range, condensation of atmospheric moisture was inevitable. Thus, the method of lubrication and the choice of the lubricant presented a considerable problem. The solution was a special lubricant, and

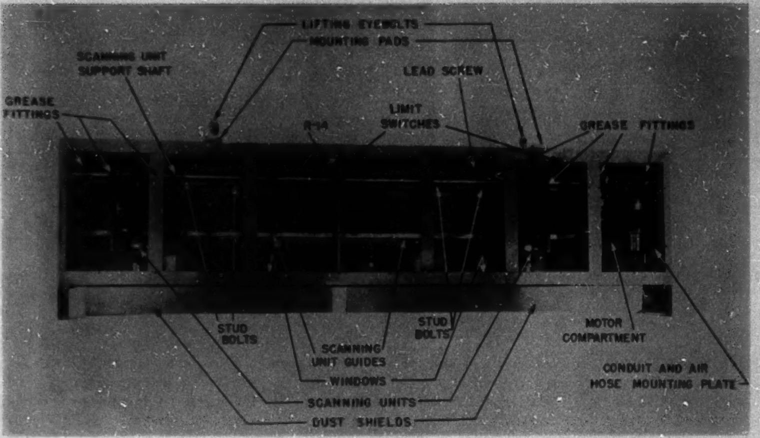

LIFTING EYEBOLTS
MOUNTING PADS
SCANNING UNIT SUPPORT SHAFT
LEAD SCREW
GREASE FITTINGS
R-14
LIMIT SWITCHES
GREASE FITTINGS
STUD BOLTS
SCANNING UNIT GUIDES
STUD BOLTS
MOTOR COMPARTMENT
CONDUIT AND AIR HOSE MOUNTING PLATE
WINDOWS
SCANNING UNITS
DUST SHIELDS

Fig. 5 (left). Detector head for width gauge (covers removed)

a method of keeping the surface properly covered at all times had to be found. The assembly in question is shown in Fig. 6. Had this problem been "solved" by inserting a note in the Instruction Book to the effect that a regular daily schedule of lubrication would be necessary, the equipment would have been subject to more down time than would have been acceptable, and it would ultimately have been removed from the production line because rust and corrosion would have ruined the ways, after which failure to give the desired service would have occurred.

Lubrication is frequently a design problem because of the choice of proper lubricant, or because of the necessity to find a simple way to provide for its renewal. It may be necessary to drive a potentiometer or a mechanical counter at a speed which requires gears, and it is astounding how frequently the designer makes no provision for lubrication other than to expect the user to put a little oil or grease on the gears occasionally. A part which is hard to reach in order to renew the lubricant will be neglected. This is too frequently the condition met in industrial design, so in any list of the problems such a designer must face, the problem of lubrication cannot be neglected if reliability is important. It is fortunate indeed that in this field it is not often that the temperature range must be as wide as it is in certain military equipment.

Heat introduces another problem in industrial design and must be considered in the design of reliable equipment. Of course, most designers attempt to keep temperatures within acceptable limits; it is normal to locate components so as to avoid causing overheating of other parts. This is not limited to the industrial field, but there are some environmental conditions which pose special problems from a reliability standpoint. Many industrial processes require large amounts of heat, and hence equipment used in such industries meets different ambient conditions than those found in other fields. Components are not always available which can withstand direct exposure, and simple means for heat removal may not always

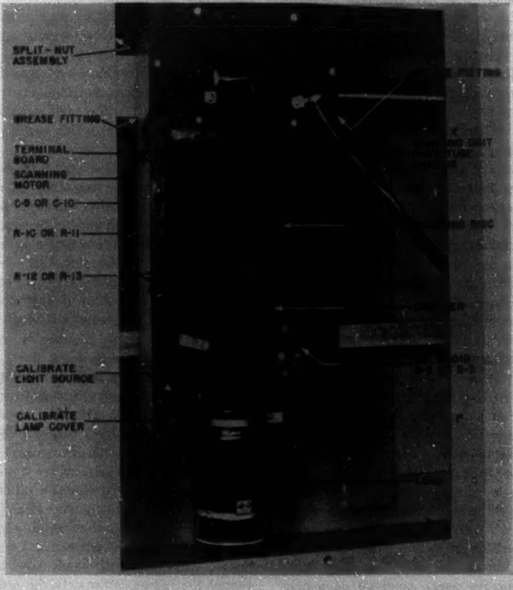

SPLIT-NUT ASSEMBLY
GREASE FITTING
TERMINAL BOARD
SCANNING MOTOR
C-9 OR C-10
R-10 OR R-11
R-12 OR R-13
CALIBRATE LIGHT SOURCE
CALIBRATE LAMP COVER

Fig. 6. Scanning unit for width gauge

be possible. Photoelectric cells, oil-filled capacitors, resistors whose values are critical (as in measurement circuits), or even transformers of ordinary design do not fulfill the reliability requirements at elevated temperatures. Yet the economics of today forces the use of electronic equipment in many instances simply because other means are not adequate or do not exist. It is not infrequent, therefore, that auxiliary cooling means must be employed. Here the designer meets two problems. These are the necessity for providing:

1. Simple and easy access for maintenance testing purposes;
2. Protective devices to announce any failure of the cooling means.

If the first is not solved, maintenance gets neglected and ultimate failure of the equipment occurs. A proper solution for the second is difficult, because some of these protective devices are reliable enough for certain fields but not for others, and a wide range of choice in such devices does not exist.

Since convenience in servicing frequently determines if not only the desirable but also the necessary things can be done in the available time, the design chosen assumes great importance in achieving service reliability. Industrial applications cannot always achieve the frequency of service testing or correction of operating deterioration which may be desirable; hence, design to obtain greater freedom from the need of frequent adjustment or a long life in a component must not be lightly side stepped by a neat phrase in an instruction book which, in effect, transfers the problem to the user. Electronic devices are prone to alter their performance due to a change of such matters as the ambient temperature or moisture, the voltage or frequency of the primary power supply, or a change of tubes. Circuits can ordinarily be designed to overcome or at least minimize many such defects, yet it is surprising how often such factors are overlooked in both the development and design stages, with the result that the equipment can be tolerated in the industrial field only if no alternative exists. A measurement instrument which requires frequent readjustment to retain its calibration, regardless of however cleverly conceived, excites far different feelings on the part of the industrial user than it did in the case of its development engineer. Cleverness of concept is of importance only if it serves without the need for greater skill or attention on the part of the user.

The statement that satisfactory reliability depends upon good design, and that good design includes proper choice of components will be readily accepted. But what is not so simple is to decide in advance what constitutes a satisfactory design or component. This becomes especially true of a new line where only a few equipments are to be built. It becomes even more difficult where the parts are purchased, and no knowledge or control can be had of the manufacturer's processes. Figs. 7 and 8 show a piece of apparatus designed as a developmental sample with what most engineers would have regarded as satisfactory components, but the design shown in Figs. 9 and 10 reveals what improvements were made in this regard when field experience, obtained from the developmental sample, was incorporated into the production design. The problem becomes more difficult when this course of events is not possible. Fig. 11 shows one part of the same equipment of Fig. 9 that endured considerable mistreatment and still was found to be in working condition when received at the factory for replacement. The problem in industrial design is to choose components which stand up under service conditions such as this equipment has endured, and then to use them in a way to favor their functioning under field conditions. It is a disheartening experience to have an expensive structure fail merely because a limit switch had been chosen which had been used to prevent excessive travel of some part, and had been designed so its mounting holes permitted reassembly in a reverse manner by a service man, with the result that it did not function at the proper point in the motion of the part to be protected.

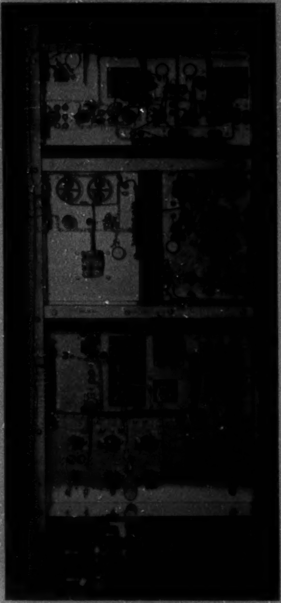


Fig. 8. Electronic cabinet for developmental sample of X-ray thickness gauge for hot-strip steel mill

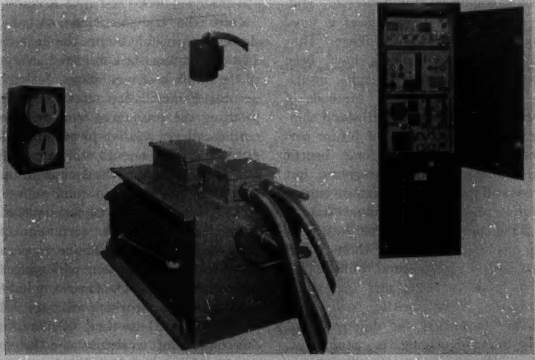


Fig. 7. Developmental sample of X-ray thickness gauge for hot-strip steel mill

In this connection, it is well to avoid moving contacts, as well as to use components well under rating. Greater freedom from motor troubles will be experienced if 3-phase induction motors are used rather than single-phase motors which have been more usual in industrial work,

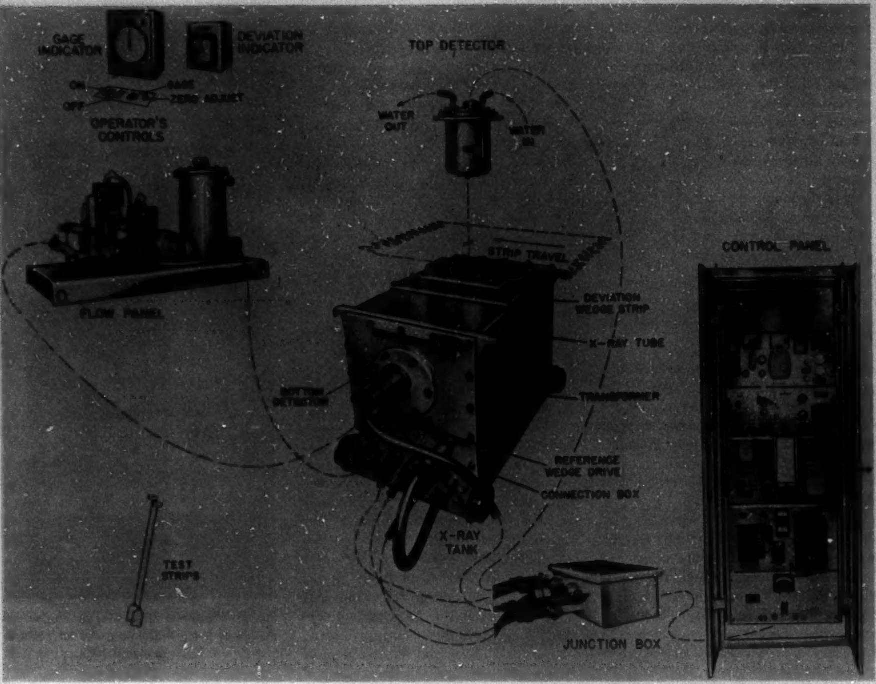


Fig. 9. Product sample of X-ray thickness gauge for hot-strip steel mill

merely because the starting switches of single-phase motors have been known to refuse to function under certain conditions of atmosphere, especially if corrosive fumes and moisture are present. Fig. 12 shows just such an application, where it was found desirable to use the 3-phase motor, and at low loading.

Connection wiring, for example, is now made with thermoplastic insulation and, under controlled conditions, it is excellent. It also has many superior properties, one of which is important where fungus is involved, but will it serve in industrial life where the combination of all operating conditions, and especially the temperature, are not exactly known to the equipment designer? Tube sockets of various designs are available, but what happens to their insulation and contacts under normal industrial use, as well as when accidents cause overheating? The literature available on purchased parts can hardly be said to be fully instructive to the equipment designer.

The designers of the components, as well as the equipment, ordinarily attempt to forsee many of the field problems, but they do not always know how the transporting agency will handle the equipment and hence may not design it for the handling conditions it will receive. Hidden damage from these causes is more serious than obvious breakage. How many transformer failures have been caused by laminations damaging insulation due to such shocks? Methods of securing parts which will withstand shipping shocks without loss of future performance have received some benefit from military electronic programs, but there is still much room for improvement. Vibration and shock in summer time is one thing, but in the icy cold of winter, parts made of rubber, or those which would receive some natural resilient protection from varnishes and such compounds, may become brittle; such conditions are frequently far different, and much more damaging to equipment. Nor can the customer wait till summer provides the conditions which may give some additional measure of protection in this manner from the rough handling involved.

Installation Aspects of Reliability

Once the equipment arrives at its destination, installation is the next step. Installation must be considered as a part of the problem of reliability. Instead of an eager installation crew, anxiously awaiting the equipment which previous correspondence relative to prompt delivery might lead one to suspect, it frequently happens that the equipment must stand around, often in rain, snow, or dirt, and be hit, moved from place to place, and sometimes even damaged before circumstances combine to permit installation. Design and packaging for such conditions must be forseen, or hidden damage may result which will not exhibit itself until some later time. Failure from this cause is just as serious as if the components themselves had failed for other reasons.

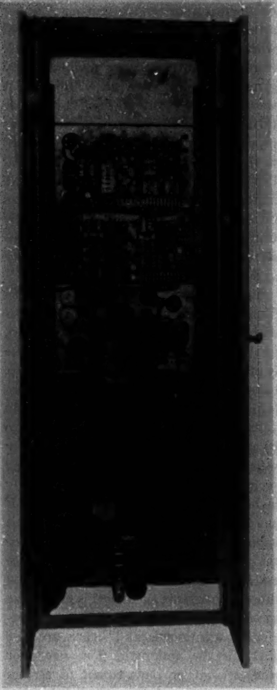


Fig. 10. Electronic cabinet for product sample of X-ray thickness gauge for hot-strip steel mill

There are other problems involved in reliable operation which arise from installation. Exposure to electrical noise is a problem that the circuit engineer is supposed to understand very well, while the mechanical engineer is supposed to understand the problems introduced by vibration. Proper operation in spite of such problems must be forseen before concrete foundations, rooms, or expensive facilities are provided to receive the equipment; otherwise, marginal utility and customer dissatisfaction are likely to result. This is part of the necessary application engineering. An equipment used to control industrial processes has to be engineered in its installation just as surely as it does in development and design. An operator's pulpit frequently requires certain location of instruments, indicators, and control accessories which may demand connection leads of such length or cause an electrical exposure which the designer did not forsee, and hence for which provision was not made. Low signal levels, poor signal-to-noise ratios, and excessive control-circuit resistance can be the reason for unreliable operation, as the other operating facts combine in varying ways from day to day.

Such a simple thing as grounding circuits can often introduce future problems which affect accuracy, and even the life of components. Examples can be cited from experience where heavy machine currents flowed through delicate electronic equipment to its detriment, merely because the ground on the electronic equipment had a lower resistance than that on the mill equipment. In such cases, the need for adequate application engineering to insure successful, as well as reliable, performance can be painfully evident. Heavy power currents in the ground system of an electronic equipment can introduce electrical noise of an amount which would seriously alter the reliable performance otherwise available from the equipment.

Another installation problem that has affected reliability is concerned with ambient conditions. Certain industries do not have clean atmospheres for electronic equipment in their mill, but often can provide such conditions only in special rooms, such as motor rooms, or out of doors. Cases have been experienced where the electronic equipment was located in a space where the ambient temperature at certain periods was such as to allow maximum hot-spot temperatures in components to approach, and even to exceed their normal temperature limits. It is one thing to meet average ambient conditions in the design, but quite another to be sure that the ambient conditions never exceed the limits set by reliable design. Long periods of abnormal ambient conditions of temperature or moisture can be as fatal to good performance as poor design would be. In such cases where this problem can be forseen in advance of design, it can be solved by using components below their normal power ratings or by applying proper protection. In cases where the design cannot be changed to accommodate what amounts to a special circumstance, special solutions must be applied. This can be the case for temperature, humidity, and fumes. Moreover, temperature cycling is also a known cause of drift in characteristics, and even in equipment failure. The point being stressed here is that the installation cannot be overlooked in any consideration of reliable performance of equipment.

Maintenance and Service Aspects of Reliability

Two final problems which must be solved in providing the customer with satisfactory performance of his equipment are those of maintenance and service. All things change to some extent; electronic equipment, at least with presently available component parts, only has a definite time period before either adjustment or repairs is needed. This is especially noticeable in equipment which performs a measurement function. Keeping equipment in proper operation is directly determined by the maintenance and service programs followed. The determination of what has deteriorated will be revealed by a regular maintenance schedule and a log of the results, while the proper repair of worn or damaged component parts will be determined by the adequacy of the service.

Experience has shown that not all customers are willing to use manpower on regular preventive maintenance schedules for electronic equipment. When, however, the problem of down time is serious enough, and electronic equipment is a vital solution to the production problem, the customer will provide for such a program if he understands its necessity. Much effort has been expended in getting greater life expectancy in component parts, but unless much greater progress is made in this regard, especially in vacuum tubes, etc., two concomitant

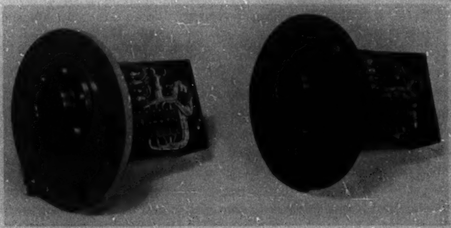


Fig. 11. Detector chassis for X-ray thickness gauge for hot-strip steel mill, showing comparison of new unit and unit accidentally subjected to extreme temperature conditions

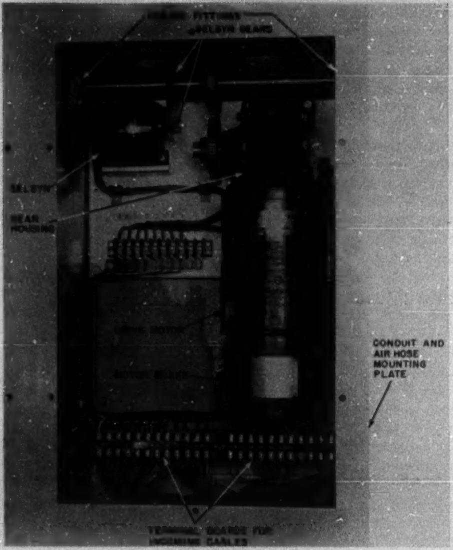

GREASE FITTINGS
SELSYN GEARS
SELSYN
GEAR HOUSING
CONDUIT AND AIR HOSE MOUNTING PLATE
TERMINAL BOARDS FOR INCOMING CABLES

Fig. 12. Three-phase motor drive for cameras in detector head of width gauge for hot-strip steel mill

things are inevitable:

1. The application of electronic devices to industry will be greatest in essential fields and where no other methods serve adequately;
2. Regular preventive maintenance programs will be unavoidable, if necessary down time is to be avoided.

It has been demonstrated that where a maintenance log is kept on a regular basis, down time has been reduced. Table I shows a limited amount of data, extracted from one such log taken from a customer's records. The results of only two tubes in one small part of the equipment are listed; similar data for other parts of the equipment would serve no additional purpose here and hence was not included, nor were results on other test points.

In the case of vacuum tubes, a replacement program based upon percentage deterioration from installation value in some significant factor, such as mutual conductance or anode current, under some special condition, was employed with success, as shown in Table I. Here, the attempt to maintain any given value of the factor, as long as it was adequate, was of less importance then the deterioration in use. This procedure is one suggested approach to the problem of increasing system reliability by an attempt to find tubes where impending failure is likely. It arises because it is known that the present tube testers are not entirely reliable in themselves in predicting if a tube is satisfactory at any given time.

In replacing vacuum tubes, it has been found desirable to avoid using new tubes without some previous burn-in period, in all cases where reduction of down time is of great importance. Other programs have been suggested, such as first establishing some heater voltage or supply voltage at which vacuum tubes known to be satisfactory fail to provide completely satisfactory operation, and then dropping such voltage to some value slightly higher than the limit so established. If any failure in performance is found at this value, the vacuum tubes which caused the failure of equipment performance are supposed to be replaced with good tubes. Programs such as this are not as acceptable to the user as the means previously mentioned, and hence are less likely to be followed. The need for readjustment or replacement of certain components has been revealed by the changes or drifts that occur from a previously established voltage of current reading at preselected points in the circuit or, in some special cases, by wave shapes of certain specified signals at designated measurement points. Here

Table I. Maintenance Log for V-501 and V-502 Tubes

1. Tube: V-501(2X2A)—900-volt rectifier
Location: Power supply
Installed with gauge; reading on tube tester = 140+140 on 5/18/49

85 on 4/1/50 85 on 4/21/50 85 on 5/5/50	115 on 5/31/50	140 on 2/18/51	108 on 9/27/51	120 on 3/28/52
	↓ 15 test periods	↓ 7 test periods	↓ 7 test periods	↓ 13 test periods
82 on 5/18/50	67 on 1/18/51	105 on 9/24/51	99 on 3/14/52	115 on 12/12/52
discarded on 5/31/50	discarded on 2/18/51	discarded on 9/27/51	discarded on 3/28/52	still in operation

2. Tube: V-502(6SJ7)—voltage-regulator
Location: Power supply
Installed with gauge on 5/18/49; reading on mutual-conductance meter = 82

82 on 5/18/49	96 (new tube)	85 (new tube)	95 (new tube)	89 (new tube)
97 67 69 67 68 64 65 ↓ 8 test periods	↓ 21 test periods	↓ 8 test periods	↓ 18 test periods	↓
42	75	78 on 9/24/51	80	84 on 12/12/52
discarded at next test period	discarded at next test period	discarded at next test period	discarded at next test period	still in operation

again, a log of past results will be found most useful. This procedure may sometimes employ, as the standard for such purposes, the average readings obtained from a number of examples for that line of equipment; but, because of the manufacturing tolerances employed in the value of components, it is more likely to require the determination of a set of such readings on the individual piece of equipment in question. The latter makes problems for the instruction book and also adds to the problems of special service from the home office. It is, however, more representative of the customer's equipment.

Training of Personnel

There is an additional factor of importance in the problem of reliability of equipment in the field. This is the training of the personnel used in each of the seven distinct groups which are involved with the equipment itself. These groups are application engineering, development, design, test, installation, maintenance, and service.

It is indeed difficult enough for either the development or design personnel to intimately know the operating conditions and requirements of even one branch of industry, and if various kinds of equipment must be applied to various industries, the services of an adequate application engineering group for each industry are necessary. In small organizations this function may have to be performed by some other group such as the designers who then must take the time to determine the requirements of the industry involved. In any case, the equipment will serve more reliably if it is designed for the specific operating needs of its application.

It would amaze the uninitiated to find how difficult it is for even experienced personnel to decide upon a set of specifications for an equipment to serve an industry in which they are presumably expert. Plants vary in layout and methods to an extent that sometimes results in several design modifications to meet various local conditions. Factors such as number of operating shifts, union rules, whether it is customary to maintain building temperatures over nonoperating weekend periods, the industrial habits of those who must do the installation work, the nature of the wastes which affect the atmosphere in which the equipment must work, the complexity of the required operation to be performed, and a host of often unsuspected matters known only to experts in that field can and do influence the design necessary for any given degree of reliability. It is not surprising, therefore, that the training of those who must do the application engineering is important if maximum reliability is to be achieved in the equipment.

The effect of the training and ideals of the development, design, and test groups are more obvious in this regard, depending upon the complexity of the equipment, and perhaps require only a brief mention. There is, however, one aspect that is often overlooked. The objectives of one manufacturing industry are not always those of another and, unless this fact is fully appreciated in its effects some difficulties are likely to be encountered. The wiring, mounting, and, in fact the design of the components which are customary in certain branches of electronics are not feasible in many industrial applications. This calls for special training of at least the development and design groups to provide the needed characteristics. Such matters as the ambient conditions affect and are responsible for design differences. The effect of certain kinds of dust and moisture conditions which may exist at a broadcast transmitter differ from those in a paper plant or steel mill, and such factors must be designed into the equipment. Hence, the enclosing cabinets acceptable to one branch may not be acceptable to another. Manufacturing and test procedures and practices which are practical for radio-receiver production are not equally so for industrial application. Even the personnel ordinarily employed in radio-receiver testing could not be interchanged with those employed in the testing of industrial equipment without special training. In some cases, even a different class of personnel is required. For example, the test personnel on heavy industrial equipment must often have a high degree of technical training, especially if only a few equipments of any one kind are to be produced, otherwise, very elaborate test instructions must be prepared, and the work intimately supervised by engineering talent.

In the case of installation groups for industrial electronic equipment, the problem of training depends upon the complexity of the equipment and its operating requirements. But either highly detailed plans for the installation must be prepared, and adequate supervision provided, or special training of the installation crew is required. The requirements of electronics are admittedly not those of the electric power field, and those not trained in the respective field of either could hardly be expected to provide the special requirements of the other without supervision, plans, or training. Since many matters are left to the discretion of those responsible for installation, their training can be a matter which affects the reliable performance of the equipment. It would not be difficult to forsee ways to lessen the reliability of performance of equipment by failure to observe these facts. For example, a previously installed equipment which may be operating satisfactorily can be affected by the installation of additional equipment in its vicinity. Regulation of power supply lines is a notorious factor in its effects upon electronic equipment. Equipment with photoelectric devices can suddenly cease to provide satisfactory service upon the installation of welding equipment, with its intermittent flashes of light. Operation of equipment with search coils, such as metal detectors, has been ruined by the use of portable welding equipment, which was designed to employ ground return of the welding currents through the metal frame of the building, even though such equipment was used in remote areas.

The last two groups, namely, the maintenance and service groups, which may or may not be the same personnel, also have readily understood training needs. In this case, the experience of one large manufacturing concern involved in the application of electronic equipment in heavy industry has been that these groups are best trained on their own equipment at the manufacturer's plant. Others have found it possible to sell such service to their customers. However, this plan is less desirable from the customer's standpoint than having his own personnel (if these can be adequately trained), simply because in case of an emergency they can be more readily available. A recently observed circumstance in a customer's plant may be of interest in this connection: The production force of the manufacturer was paid on the basis of quantity of product produced which met specifications. They found their earnings increased following the application of a certain electronic equipment, and hence were very insistent that this equipment be in operating condition at all times. The plant management therefore found it desirable to issue instructions to the maintenance and service group to supervise the condition of the electronic equipment religiously, and to drop all other work not having special priority in case of emergency failure of that equipment. It is not difficult to understand that these men were specifically trained in their work. The emphasis placed upon such matters and how various companies achieve their objectives is revealed by the interesting practice of certain companies in applying a large hourly charge in their financial accounts against the maintenance and service groups for any down time.

Naturally, the emphasis which must be placed on the various aspects of the problem of reliability in industry applications differs from industry to industry and, indeed, differs between equipments of varying degrees of importance in any one industry, just as the emphasis will not be the same among manufacturers of similar equipment. Moreover, the degree of reliability demanded will not be the same in all fields of electronics, nor will all factors receive equal consideration. But inasmuch as it cannot be claimed that the components from which such equipment is produced are entirely reliable, emphasis on these factors is justified by the extent of their contribution to improved reliability in the performance of the final product in the customer's plant; this, after all, is the objective sought.

Reference

1. The Component Problem In Industrial Electronics, Ellsworth D. Cook. 1953 Components Symposium, Pasadena, Calif., April 29–May 1, 1953.

No Discussion

The New Jersey Turnpike—A Unique Highway Communication System

P. F. GODLEY
NONMEMBER AIEE

J. R. NEUBAUER
ASSOCIATE MEMBER AIEE

D. R. MARSH
NONMEMBER AIEE

THE NEW JERSEY Turnpike carries by far the greatest traffic load of any toll highway, anywhere. Its peak day totalled 84,132 vehicles; its peak month, 2,004,000; for the year 1952, 18,239,000. Its cost will total some $285,000,000; that of its communications system and related facilities, less than 0.1 per cent of this figure.

The Turnpike now reaches from Deepwater, N. J., on the Delaware, to the George Washington Bridge across the Hudson. Extension is expected for connection with the New York State Thruway now in construction. Its length will then be about 130 miles. Its radio system gives coverage for mobile units over this entire distance, plus telephone and teletype interconnection of its three police stations with the State Police Headqurters at West Trenton.

The system embraces:

1. Five cochannel very-high-frequency (VHF) stations, operating in the 150-megacycle band, to mobile units from discrete locations;
2. Seven 950-megacycle microwave stations, which are the backbone of the system;
3. Some 75 mobile units of the police, maintenance, emergency service, and supervisory groups;
4. Some 86 toll booths at the present 17 interchange stations.

Some of the salient factors entering into the development of specifications for the system were somewhat as follows:

1. An analysis of probable traffic load requirements. This included consideration of the estimated activities of toll, general maintenance, police, and supervisory groups, and emergency services for patrons;

2. An analysis of allocation possibilities for VHF channels covering the length of the Turnpike, as planned and projected. Here, the advantages versus the disadvantages of operation in the 150-megacycle band, as against the 450-megacycle band, were weighed. Propagation characteristics of the 150-megacycle band were superior, but channel availabilities were extremely few. On the other hand, commercial development in the 450-megacycle band had been limited.

Throughout the study, attention was held centered upon spectrum economy, which, within New Jersey, was a prime essential in view of proximities to the great New York and Philadelphia metropolitan areas. Also, mobile units must be free to roam the full distance without necessity for frequency change, and this dictated the use of a single pair of VHF channels plus the critical evaluation of possibilities for cochannel and adjacent channel interferences.

But two suitable channels within the 150-megacycle band were available: 154.83 and 155.19 megacycles; these were used. Their separation is but 0.36 megacycle, which called for study of special means for effecting suitable isolation of transmitter and receiver antennas if economies were to be derived through the use of single towers for the support of both.

3. Topographical studies were made with attention centered upon maximum economy with respect to location of VHF relay stations at points where planned buildings could house equipment; and, in the interest of further economy, the number of relay stations was minimized. Suitable propagation tolerances were injected, and where, due to topographical features, there appeared to be merit in so doing, experimental propagation studies were carried through.

4. Economies of both spectrum and cost pointed toward the desirability of concentrating VHF radiation along the axis of the highway. Thus, a 2-element, broadside antenna using half-wave dipoles was designed, experimentally checked, and specified for both reception and transmission, these to be mounted upon the same tower at each station in an essentially coaxial relationship, and with vertical spacings which had been determined, both mathematically and experimentally, as sufficient to reduce mutual impedances to quite inconsequential values. This arrangement found the use of cavity resonators, or the like, unnecessary; and transmitter operation causes no desensitization of the corelated receivers, notwithstanding the small channel and physical separations.

5. An antenna solution of somewhat different character was applied, for the resolving of that difficulty encountered when a mobile unit finds itself in midzone between two simultaneously-operating cochannel stations. Here, even though both carry the same speech modulation, interference is developed by phase differences in the audio-frequency components of the two carriers. The receiver cannot discriminate.

A novel 2-element, directive, car-top antenna was devised, wherein one element is driven, the other parasitic. A 5-to-1 front-to-back field ratio is obtained. Reversal of the field pattern is brought about via a feed-line switching relay, actuated, in turn, by a toggle switch on the dash-mounted control plate. Thereupon the parasite is driven, and vice versa.

With, say, 25-mile spacing of relay stations, the effect of antenna pattern reversal is about the same as that had were the vehicle instantaneously lifted to a point some 4 miles up or down the highway, depending upon whether the switch is in either the "fore" or "aft" position. The interference area is, thus, erased.

This antenna also served an important economic purpose. It made possible the use

---

Paper 53-213, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted Jan. 8, 1953; made available for printing April 10, 1953.

P. F. Godley is with the Paul Godley Company, Great Notch, N. J., and J. R. Neubauer and D. R. Marsh are with the Radio Corporation of America, Camden, N. J.

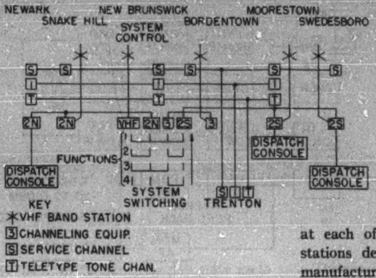


Fig. 1 (left). New Jersey Turnpike microwave-very-high-frequency system

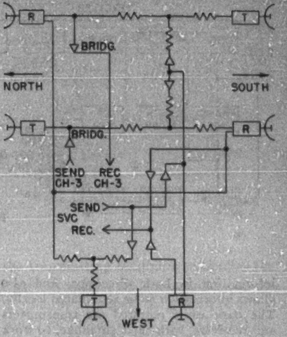


Fig. 3 (right). Bordentown leg repeater

of 10-watt mobile transmitters, which could be powered without costly amplification of standard battery-generator car equipment.

6. Because of expected heavy commuter-traffic loads in the metropolitan areas, and overall congestion over week ends and holiday periods in summer, greatest practicable flexibility of the system was deemed a prime essential. Furthermore, control should, it was held, be centered at the New Brunswick Administrative Headquarters, with auxiliary control, either for the system as a whole or pertinent sections thereof, available to the police substations at Newark and Moorestown. This dictated the tying of the 5 VHF stations together—and multichannel microwave facilities were used for this purpose. Details as to flexibilities will be covered in the technical description that follows.

Microwave equipment is accommodated at the 5 VHF stations, where the single support tower carries all antennas. The Newark police substation is contiguous to the airport there, on which account microwave terminal antennas are mounted but some 20 feet above grade level, and oriented upon the Snake Hill microwave relay 5 miles to the north. The West Trenton station is a microwave branch terminal only, although the support tower there now also accommodates other State Police radio facilities.

Mutual impedances between immediately adjacent microwave transmitting and receiving antennas are essentially nil, their respective dispositions being vertical and horizontal.

In order to fulfill the specification detailed by the Turnpike Authority, many special facilities and adaptations of standard equipment were required. Considerable effort was devoted to choosing basic components to fit a standard pattern. The diversified functions appearing at each of the terminal and repeater stations demanded custom design and manufacture of distribution and control equipment, having many interlocked and carefully coordinated circuit functions. The design and development of the Turnpike system was, therefore, divided into two basic phases, one the choice and application of standard items and the other, the analysis and design of distribution and control equipment.

Over-all System

The functions required at each of the control and administrative dispatch locations are shown in Fig. 1. It can be seen that up to five independent channels were necessary to provide the desired services. They were arranged as follows:

1. Newark to Snake Hill
S=service party line
2N=VHF control and audio

2. Newark through Snake Hill to New Brunswick
S=service party line
2N=VHF control and audio
1=intercom party-line telephone
T=party-line teletype

3. Snake Hill to New Brunswick
S=service party line
2N=VHF control and audio

4. New Brunswick to Bordentown
S=service party line
3=VHF control and audio

5. New Brunswick through Bordentown to Trenton
S=service party line
1=intercom party-line telephone
T=party-line teletype

6. New Brunswick through Bordentown to Moorestown
S=service party line
2S=VHF control and audio
1=intercom party-line telephone
T=party-line teletype

7. Trenton through Bordentown to Moorestown
S=service party line
1=intercom party-line telephone
T=party-line teletype

8. Moorestown to Swedesboro
S=service party line
2S=VHF control and audio

Essentially, the system consists of three terminals and four repeaters. All of the repeaters are of the type normally designated as a drop; however, the arrangement used for this system required some variations. The drop repeater at New Brunswick included a provision to separate the VHF control and audio, north and south, and the Bordentown drop repeater provided a leg circuit to Trenton.

For those not familiar with microwave practice, a definition of station types is in order:

1. A terminal station is located at the end of any portion of a system, which may be the extremes of the main line or a side leg circuit. The main elements are a transmitter and receiver, with their associated antennae and terminating facilities.

2. A drop repeater is located at intermediate points within the main line or leg portions, and provides a medium to pass

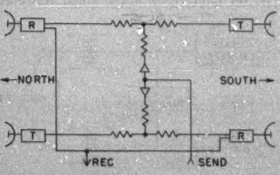


Fig. 2 (left). Snake Hill-Moorestown drop repeater

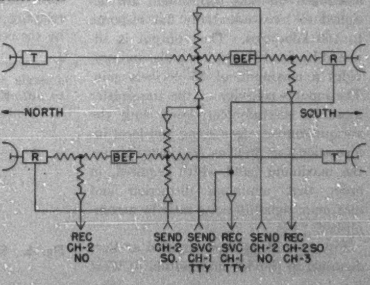


Fig. 4 (right). Modified terminal repeater

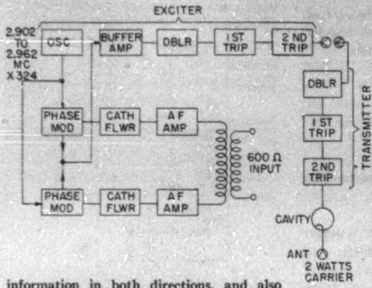


Fig. 5(A) (left). Microwave transmitter, 960 megacycles

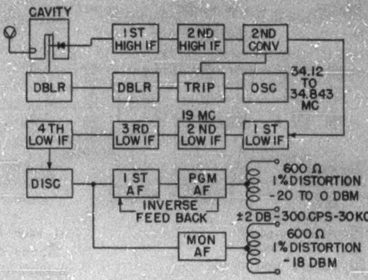


Fig. 5(B) (right). Microwave receiver, 960 megacycles

information in both directions, and also allows information to be terminated locally. The basic units required are transmitters and receivers, with their antennae, to work in two directions, and a bridging and insertion network of amplifiers.

3. A leg repeater is used to bring information from a point off the main line and insert in one or both directions. It consists of a drop repeater which is terminated in a terminal station. In some instances local information will be bridged across the send-receive line.

4. A terminal repeater is essentially a back-to-back connection of two systems at their terminal points, providing some through-information channels, and dropping or inserting others.

The repeater stations, as used on the Turnpike are detailed in Figures 2, 3, and 4. Fig. 2 is the straight drop repeater, as used at Snake Hill and Moorestown. The receiving line is provided by bridging the outputs of fixed-gain amplifiers located in the microwave receiver. These amplifiers are normally designated as a monitoring output. The through connection is accomplished by interconnecting the microwave-transmitter audio input with the receiver-program audio output. The program output is an adjustable low-distortion amplifier, also located in the microwave receiver. To insert information in both directions, a resistance tee and bridging amplifier network is connected, to provide a common outgoing audio input. The amplifiers used for this purpose are individually assembled on rack strips, and are designed to have essentially flat response to 150 kilocycles. Their output is adjusted by fixed plug-in tee pads, and provides a maximum of 35 decibels gain. The input sensitivity of the transmitter must be carefully correlated with the various sources of on-going and local information; as well as consideration for the maximum band width required, in order that minimum distortion and maximum signal-to-noise ratio be accomplished.

Fig. 3 is the leg repeater used at Bordentown to provide information to West Trenton State Police headquarters, and through information to the Moorestown-Swedesboro section. Basically this arrangement follows the pattern of the drop repeater, using the drop circuit to connect the leg circuit to the main system. The local service channel and terminating facilities are bridged onto the main sending and receiving circuits, using the broad-band amplifiers for isolation and level control.

Fig. 4 is a modified terminal repeater as used at New Brunswick. This station is the central administrative dispatch of the entire system, and requires terminations sufficient to route or divide the VHF-control and audio information to all sections of the Turnpike. It combines the facilities of a drop repeater and a terminal repeater, by inserting band-elimination filters to divide the through circuits from the terminal circuits. To the service, interphone, and teletype channels it serves as a drop repeater, but to the VHF-control and audio circuits it appears as a back-to-back terminal allowing information to be interconnected in any arrangement desired. It can be seen that as in the other repeaters, the broad-band amplifiers are used to control the levels and provide bridging isolation.

Referring again to Fig. 1, the interconnection of the VHF circuits is arranged for four operating systems, all of which can function independently as automatic relaying systems, or can be supervised by the central dispatcher.

1. Circuit 1 interconnects all elements as a single party line, permitting any section to automatically relay through any other section, and allow the dispatcher to broadcast or communicate with any unit over the entire length of the highway.

2. Circuit 2 interconnects stations of the north portion and the south portion as separate systems, permitting each to conduct traffic as required, independent of the other.

3. Circuit 3 interconnects the north portion with the Bordentown station, and allows the Moorestown-Swedesboro section to operate as a unit.

4. Circuit 4 divides the north portion into two independent stations and connects the Bordentown station to the Moorestown-Swedesboro group.

The flexibility emphasis is placed on the north section of the highway, which obviously carries the largest volume of traffic.

There are terminating facilities at all stations, both terminal and drop repeaters; however, only the service channel is common to all. The interphone and teletype are party-line dropped at the four State Police offices, to provide high-volume routine traffic circuits. The VHF control and audio circuit appears at all stations except Trenton. Five of the stations are equipped with a VHF transmitter and receiver. The Newark terminal functions as a VHF dispatch point, using the Snake Hill VHF facilities whenever it is separated from the remainder of the system. Other VHF dispatch consoles are located at Moorestown and Swedesboro.

Microwave and Channeling Equipment

Figs. 5(A) and 5(B) are block diagrams of the *RCA CW5-B* 960-megacycle transmitter and receiver units, which were used

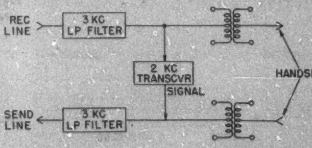


Fig. 6. Service channel audio and signalling

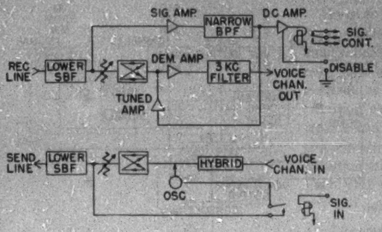


Fig. 7 (left). Single-side-band-suppressed-carrier multiplex channel

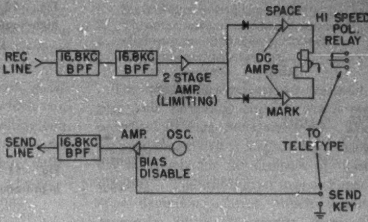


Fig. 8 (right). Teletype channel

at the terminal and repeater stations to provide the necessary audio band width to carry the multiplexed channels. The transmitter consists of a broad-band frequency-modulation and phase-modulation unit, frequency modulation being accomplished below 2,000 cycles, and phase modulation being effective up to 30 kc. The audio information is applied through a push-pull amplifier and cathode follower to the phase modulator, the cathode follower being used to lower the impedance and improve the linearity over the modulation band width. The exciter unit provides an approximate 50-megacycle output to drive the transmitter and provide a 2-watt carrier output. The crystal control-oscillator frequency is multiplied 324 times and is tuned in the final stage with a specially designed cavity which encloses the output tube.

The microwave receiver is designed to provide less than 1-per-cent distorted output over ±2-decibel 0.3- to 30-kc range at 2 output sources. The program output can be increased from −20 to 0 decibels-referred-to-1-milliwatt, by the use of inverse feedback and, therefore, has available a gain control for this purpose. The monitor output is a fixed gain stage without inverse feedback, providing −18-decibels-referred-to-1-milliwatt, output. The receiver is a double superheterodyne, using a cavity resonator and crystal diode for the first converter stage. The second conversion frequency produces a 19-megacycle low intermediate-frequency having a 0.5-megacycle band width.

The service channel, as shown in Fig. 6, was provided for the use of the maintenance and engineering groups, and used the lower 3,000 cycles of the base band. It is separated from the upper portion by low-pass filters having 40-decibels attenuation above 3 kc. Within this pass band, a 2-kc tone is transmitted to operate signalling buzzers at all the equipment locations.

The intercom telephone, VHF control, and audio channel equipment is shown in Fig. 7. These channels use the frequency-division principle of multiplex and employ the single-side band suppressed-carrier transmission method. This is a time-proved communication practice, which has proved most easily adapted to this type of operation. At the expense of fundamentals which may be familiar to some, it would be well to discuss the general characteristics of a single-side band suppressed-carrier transmission system. The voice information is applied through a hybrid coupling to a balanced-ring modulator system, consisting of a quad of varistors. At the balance point of this system, a carrier frequency is injected. This produces upper and lower side bands without a carrier. Passing the two side bands through a suitable filter, the upper side band is eliminated, and only the lower side band is transmitted to the base-band input of the microwave system.

To receive this signal, the lower side band is selected by a similar filter and again is applied to a balanced varistor ring. In a standard channel, a local oscillator tuned to the same frequency as the transmitting oscillator resupplies the carrier frequency at the input of the demodulator amplifier, the output of which would be supplied to a voice channel circuit through a 3-kc filter. Because of the necessity for low distortion and minimum oscillator-frequency drift, which is desirable when applying these signals to a VHF transmitter for retransmission, it is desirable to synchronize all receiving channels to the transmitting channel. This is accomplished by utilizing the ringing-signal circuit, which is normally used to operate a telephone-ringing generator or dial circuit. The ringing signal is normally supplied by bypassing the balanced varistor of the transmitting circuit, reinserting the carrier at the input to the lower-side-band filter.

This, then, provides a signal carrier with the lower-side-band voice frequencies, which, when received at the other end, is amplified at the output of the receiving side-band filter, and passed through a narrow band pass-filter to a d-c amplifier to operate a signalling relay. In this system, this information is applied to the balance varistor ring through a tuned amplifier, replacing the normal local oscillator. This provides a single frequency throughout the system. You will note in Fig. 7 that the d-c amplifier is also provided with a disabling circuit. This is a bias applied to cut-off the plate current, and is controlled by external relay contacts when used in conjunction with the protective lockout system. Using this method it is possible to achieve better than a 40-decibels

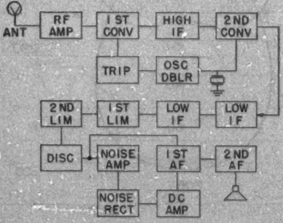


Fig. 9(A) (left). Mobile receiver

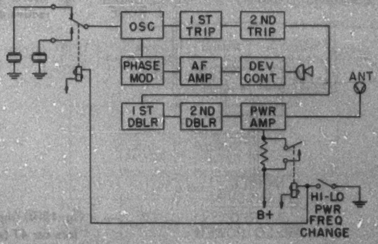


Fig. 9(B) (right). Mobile transmitter

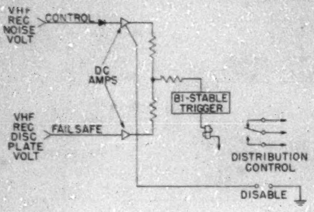


Fig. 10 (left). Very-high-frequency carrier operated relay

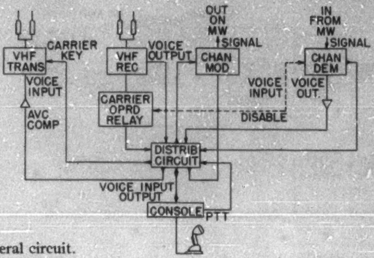


Fig. 11 (right). Very-high-frequency control system

signal-to-noise ratio at the voice frequency terminals.

The teletype circuit is shown in Fig. 8. This is a single-tone carrier, keyed by a model *15* teletype printer thru a bias-disabled amplifier and band-pass filter. The mark signal is represented by no carrier, and the space signal by carrier transmission. The signals are received through a cascaded narrow-band-pass filter by a two-stage limiting-type amplifier, the output of which is rectified to provide control bias for the space signal. The mark amplifier is protected from the space signal by an oppositely connected rectifier, to maintain its plate cut-off during the space signal. When no signal intervals occur, the normal bias causes this d-c amplifier to draw current. The two currents are applied to the opposite windings of a high-speed polarized relay, which in turn, is interconnected to the receiving loop of the printer. The tone transmitter and receiver are arranged in a half-duplex circuit, allowing the machine to receive its own signals as well as any other machine's. This arrangement is convenient for break-in and party-line operation.

VHF Station Equipment

Figs. 9(A) and 9(B) detail the basic circuits employed in the VHF base-station receivers and transmitters. The HI-LO power and frequency-change features are a part of the mobile transceivers only, which use the same general circuit. The transmitter is the phase-modulated frequency-modulation type, which multiplies the crystal frequency 36 times to produce a 50-watt carrier at 154 megacycles. The deviation control used at the stations is a separate automatic volume control-compressor-type amplifier, providing 4 decibels output variation for 30 decibels input change. This same unit is used to control interconnecting levels at New Brunswick; therefore, we will defer its description until we discuss the central dispatch facilities. The VHF receiver is a double-conversion superheterodyne using a single crystal oscillator with a multiplier system. The low intermediate-frequency stages drive 2 limiters and a discriminator to provide 20 decibels of noise quieting for less than 0.5 microvolt of signal at the antennae. The audio amplifier is muted, when no signal is present, by a rectifier and d-c amplifier which control the grid bias of the first audio-frequency tube. The preemphasis and deemphasis characteristics are designed to follow a 6-decibel per octave curve. The overall distortion is less than 5 per cent under all but maximum automatic audio-level control conditions.

The control of the system by the VHF base station receiver is effected through a carrier-operated relay, detailed in Fig. 10. This unit utilizes a rectified component of the receiver noise voltage to bias the control amplifier, the output of which is applied to one section of a common divider. The other portion of the divider is connected to a second d-c amplifier, which receives its bias control from the VHF-receiver-discriminator plate circuit. The function of these two biases is to provide a control voltage which will combine reliable operation with a fail-safe protective feature. The control voltage which is derived from the rectified noise voltage of the receiver, obtained from the discriminator output, changes in a negative direction when carrier is received. This is due to the quieting effect of the frequency-modulation receiver. Such a condition easily could be simulated falsely, when the sensitivity

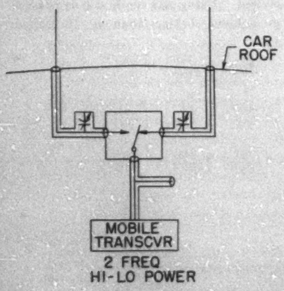


Fig. 12(A) (left). Mobile directive antenna system

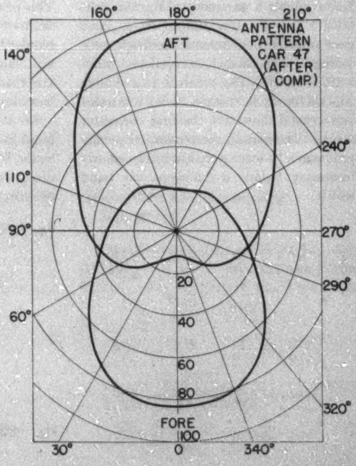


Fig. 12(B) (right). Antenna pattern car 47 (after compensation)

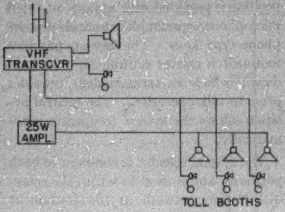


Fig. 13. Interchange system

of the receiver deteriorates to a point similar to that at which the noise voltage would decrease with carrier quieting; this would provide a false control voltage. Since the discriminator plate voltage depends upon the normal sensitivity of the receiver, it will reflect any decrease in sensitivity through the second d-c amplifier. This will shift the operating point of the divider sufficiently so that the normal control voltage will have no affect on the bistable trigger-relay control circuit.

This type of carrier-operated relay can be adjusted to a very accurate signal operating margin, and has good stability in the presence of extraneous noise in the receiver. It is very fast in its response, which is highly desirable in initiating the automatic chain of relay functions necessary to activate the carriers along the line. You will notice that a disabling circuit is also indicated on this unit. This is accomplished by supplying a large positive bias to the control circuit whenever the disabling contacts are opened. The disabling function is accomplished by interconnecting the normally made contacts of the multiplex channeling-signal relay and the carrier-operated control relay in a crisscross fashion. In other words, whichever relay operates first disables the other, enabling it to maintain preferential control during the period of transmission.

The over-all relationship of the channeling equipment and the VHF base-station transmitter-receiver and carrier-operated relay are diagrammed in Fig. 11. At each of the 5 VHF base stations, a distribution circuit is provided which controls 3 sets of audio and signalling control circuits, with associated audio distribution and impedance matching paths. Each VHF station operates as a result of any 1 of 3 sources of information:

1. The stations own VHF receiver;
2. The microwave multiplex channel output;
3. The local dispatch console.

If a carrier is received at the local VHF receiver when it is idle, the carrier-operated relay will immediately disable the incoming microwave channel and switch the output of the receiver to the input of the VHF transmitter and the modulating circuit of the outgoing microwave channel. This audio information will also appear at the remote console.

When any one receiver or local dispatcher point activates the multiplex channel at a time when the equipment is idle, the channel unit will immediately disable its local VHF receiver and carrier relay, and supply audio and signalling information to the console and the VHF transmitter. Whenever local dispatching is effected, the telephone-line d-c control circuit operates the distribution circuit to supply audio and control information to the VHF transmitter and to the outgoing microwave channel. An amplifier is located in the output of the channel demodulator to correlate the output level with the levels of the VHF receiver and console.

Interchange and Mobile Stations

All VHF communication facilities serve two groups of transceivers located on or adjacent to the highway. One group is the entrance and exit interchanges, the other the various vehicles attached to the Authority's departmental organizations. The latter group is divided into four components:

1. State Police patrol cars
2. Turnpike administration personnel
3. Highway maintenance personnel
4. Customer service stations

The mobile transceivers use transmitting and receiving circuits which we described in Figs. 9(A) and 9(B). The receiver sensitivity is slightly less, being 0.8 microvolt for 20 decibels of noise quieting. The transmitter uses a clipper-type deviation control, which instantaneously limits the carrier swing to ±15 kc. The hi-lo power change and frequency change is used for limited range car-to-car communication. Since this frequency is the same as the base-station transmitter, only one receiving frequency is required. This has the additional advantage of leaving the car open to receive messages from the dispatchers at all times, and relieves the automatic-relaying circuits of considerable traffic.

In a system such as this, where 5 transmitters are radiating simultaneously, there will be, of necessity, overlapping areas between any two base stations. At a point where signal levels are approximately equal, a strong zone of interference will be produced, due to the phase differences in the audio components of the two carriers and the inability of the receiver to eliminate either carrier. To overcome this, a special mobile directive antenna system was devised by the Paul Godley Company, which provided a selective directional pattern under the control of the vehicle operator. The basic components of this system are shown in Fig. 12(A). The two antennae are located slightly less than a quarter wave apart, and are phased by means of the electrical length of the transmission line to utilize the undriven antenna as a parasitic director, providing a 5-to-1 field strength ratio. To match the impedance effect of this parasitic element, a stub is inserted in the transmission line to the mobile transceiver. The electrical length of the antenna transmission line is varied by means of the shunt reactive element located at the coaxial switch. This antenna system is equally effective on both transmission and reception. A typical propagation

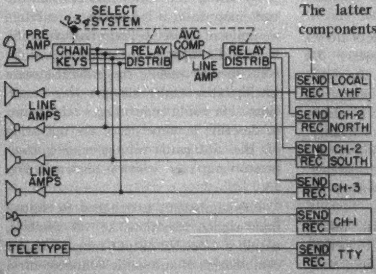
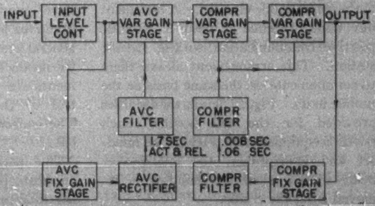
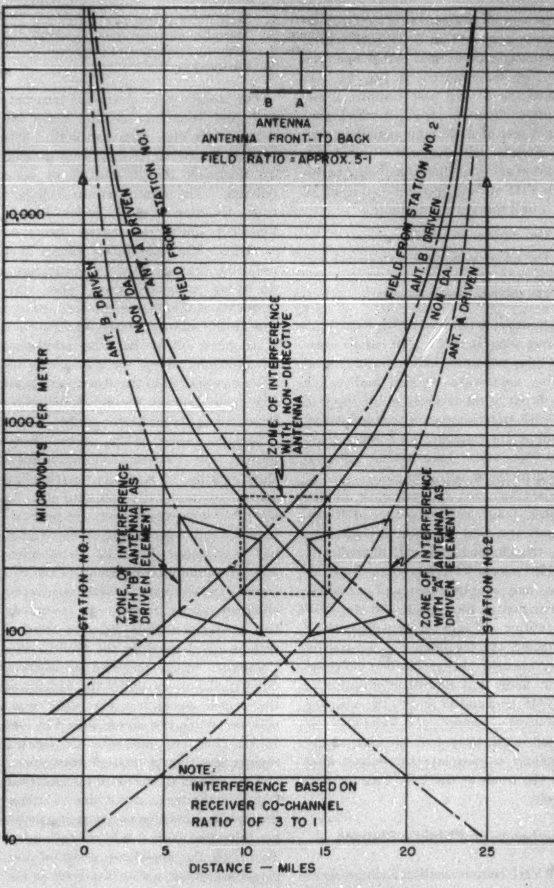


Fig. 14 (left). New Brunswick console

Fig. 15 (right). Automatic-volume-control compressor





NOTE: INTERFERENCE BASED ON RECEIVER CO-CHANNEL RATIO OF 3 TO 1

Fig. 16. Car-top mobile directive antenna

pattern is displayed in Fig. 12(B).

The interchange transceivers are a-c operated low-power units, having 15 watts carrier on the same frequency as the regular mobile channel. A 3-element directional antenna is used to provide positive communication with a specific station. This arrangement allows them to communicate on the same basis as the mobile units. Fig. 13 details the facilities provided for the toll booths. Each booth is equipped with a telephone handset and loudspeaker, the loudspeaker being muted when the handset is removed from the cradle.

Control and Dispatch

The principal operating circuits are set up by the New Brunswick operator. We have previously touched on the basic functions he establishes in the section on general system facilities. Fig. 14 shows the means by which the 4 system arrangements were brought under the direct control of the administrative dispatcher. The console facilities are shown on the left.

One position only is shown, but a duplicate position is provided to permit handling heavy emergency traffic. Each position is provided with a boom-mounted microphone, preamplifier, and 4 telephone-type keys. The key-up provides foot-switch carrier control, and the key-down, which is springloaded, provides press-to-talk operation. The neutral position allows the system to operate automatically.

The system selector is common to both positions, and is used to set up the relay-distribution network at the equipment racks for the system arrangement desired. The relay-distribution unit sets up the conditions which will provide paths between the receiving and sending circuits to reroute the interconnection information to the proper outgoing channels. In conjunction with this, the console keys allow the microphone to be directly inserted into the chosen circuit, giving the dispatcher control over the circuit, in preference to the automatic facility.

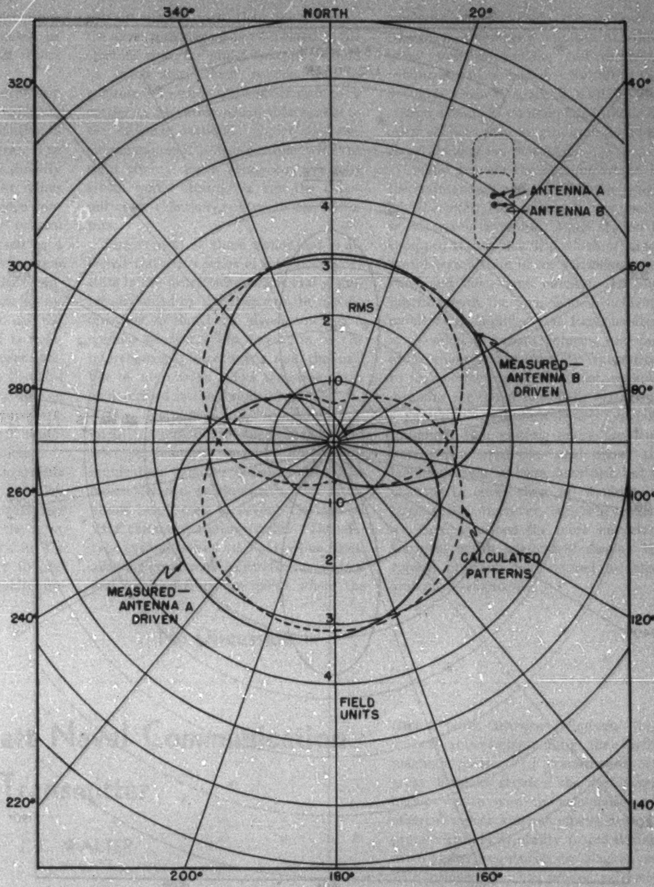
The console keys are so interlocked that the two positions cannot be operated simultaneously, but are available on a first-come first-served basis by either dispatcher. Busy indicators are provided to alert the dispatcher as to the channels which are in use. Each channel, including the local VHF station, is terminated in a separate speaker. The output of the channel is not normally connected to the monitoring speaker and amplifier, except when a signal is coming in, this operation being effected by the use of the incoming signalling circuits actuating a muting-relay arrangement.

You will note between the two groups of relay distribution that an automatic-volume-control compressor amplifier and line amplifier are used to regulate the inter-channel levels. The diagram as shown in this figure shows only one position of the system switch, each of the other three positions being a repetition of this, with a different distribution of the send and receive lines. This calls for 3 sets of automatic-volume-control compressors and line amplifiers to control the outgoing levels under all bridging conditions.

By using the automatic-volume-control compressor, it is possible to maintain a uniform 0-decibel-referred-to-1-milliwatt audio input to the channel modulators, thereby raising the general system intelligibility under all operating conditions. It would be in order, at this point, to describe in some detail the operation of the automatic-volume-control compressor amplifier, which is shown in Fig. 15.

This equipment is designed to control input-signal variations of 30 decibels, within a 4-decibel output range. It is a combination of automatic volume control

Fig. 17. Car-top mobile directive antenna, horizontal plane pattern



with a relatively slow time factor, and a fast-acting compressor to limit the instantaneous audio excursions. The incoming audio is amplified by a fixed-gain stage rectified, and applied through a resistance-capacitance time-constant filter to a variable-gain stage. This acts to provide a relatively constant signal level to the compressor input, limiting the maximum excursions, but, at the same time, maintaining the output level regardless of decreasing input levels. The operation of the compressor portion is controlled by a sample of the output information through a fixed-gain stage and rectification, which is applied, in turn, simultaneously to the 2 variable-gain stages, the attack time being 8 milliseconds, and the releasing time 60 milliseconds. The frequency response of this type of amplifier is flat within 1 decibel from 300 to 7,000 cycles, which is more than adequate to handle the normal 300- to 3,000-cycle range employed in this type of system. This results in a steady audio-level output from the base station, which increases the intelligibility in the moving vehicles under high ambient noise-level conditions.

There are, as previously mentioned, three other dispatching points, which are equipped to control the system at the discretion of the administrative dispatcher, but in preference to the mobile and interchange stations. These units are equipped with a microphone and amplifier, which is switched by the press-to-talk circuit to serve as a monitoring or talking facility. A single metallic pair is used to provide the audio path and d-c control circuit. The d-c control circuit operates from line to line, using telephone repeater-coil winding centers as terminals for operating the relay coil on the distribution unit.

Summary

To summarize the practical operation and application of this system, we can describe 4 major types of system operation which are accomplished:

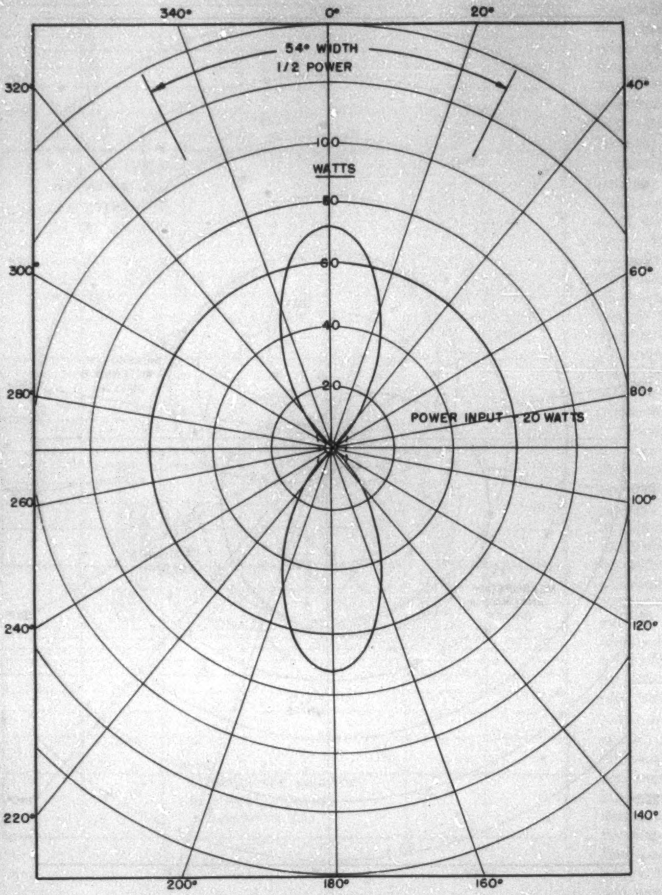


Fig. 18. Very-high-frequency base-station directive antenna, horizontal plane radiation

1. Mobile-to-mobile relay;
2. Interchange-to-interchange relay;
3. Mobile-to-interchange-to-mobile relay;
4. Point-to-point communications.

The first three of these operations are completely automatic, and provide flexible communications along the entire length of the Turnpike. A description of the sequence of operations will help to visualize the complex paths and control operations necessary, and the means by which the facilities which have been described are employed. Assuming that a mobile station wishes to initiate a call to another mobile station located at some other point on the Pike, he will activate his carrier with his press-to-talk switch, and with his antenna-directional switch positioned for the nearest station. This is easily determined, since the Turnpike locations are known by the mile markers. The mobile carrier signal will quiet the VHF base-station receiver, thereby activating the carrier-operated relay. Two things occur at this point, the first being that a ground circuit opened by the carrier relay in the multiplex-channeling equipment allows a disabling bias to be applied to its receiving-signal amplifier. This provides protection for the carrier-relay control sequence for this operation, and at the same time, disables the channel output to the distribution system. The second thing is the initiation of the control sequence in the distribution system, to activate the VHF transmitter carrier and enable the channel modulating and signalling circuit. Simultaneously, the audio output of the receiver will be applied to a dividing network, making audio information available to the local VHF transmitter and to the channel modulator. If a local dispatch control unit is tied in at this particular station, it also will receive the same information. This means that the information on the mobile carrier is now simultaneously distributed locally and at the other 4 VHF stations. The mobile unit which it was desired to contact will hear this signal, as it is made available from the distant transmitter or the microwave channeling system, or in some cases, he will be close enough to hear it direct from the local VHF transmitter.

The microwave signalling information will have also supplied disabling information and audio distribution information to the other 4 receivers in the system. The disabling information is applied directly to the other VHF station carrier relays by means of the normally grounded contacts of the channel-signalling control relay. If the system is connected as a party line arrangement, which is position 1 of the administrative dispatcher's control console, the control and audio information will drop to the audio and d-c circuits at New Brunswick and be reinserted through a relay interconnect facility automatically. As high as 7 relays in automatic sequence will be in use on signals passing through New Brunswick. The New Brunswick operator will simultaneously monitor any transmission.

Because of the channeling arrangements requiring this large number of relay operations, considerable attention had to be devoted to determining the relay time constants and the delays in the transfer of information in order to prevent system lockup. It is equally important that no more than one source of the same audio signal be connected to the channel, since the phase distortion would be intolerable. Such a condition is possible if a mobile transmitter radiates a carrier at a location which is favorable to two adjacent stations. Under the same circumstances it is also possible, with the right timing, for a relay race to start which would alternately key the transmitters without any control by the operators.

The control of these conditions is effected through a series of interlocked circuits in the distribution units, and is materially aided by the ability of the mobile operator to direct his signal towards a specific receiver. The operation of the interchange-to-interchange and the mobile-to-interchange follow the same paths as described in the mobile-to-mobile.

The point-to-point communication facilities, consisting of the telephone and teletype channels, provide direct communication facilities between all of the police division headquarters, the interphone being used to relieve the normal VHF channel of routine traffic. The teletype system is used on a party-line basis, allowing group messages to be transmitted and received from any point where the machines are provided, and is flexible enough to allow any number of additional tone-transmitting and receiving equipments to be added. Not all of the available channel space is used in this system and it is contemplated that, as other services may be required, an additional channel can be added.

Finally, but not the least important, is the maintenance and test facilities which have been provided to permit easy analysis of system troubles. Each station is equipped with a jack field which provides 4-wire termination of all pertinent base-band and voice-band circuits. In conjunction with the jack field, a bridging monitoring amplifier and local handset is provided to permit entering any one of the channels or local VHF transmitter to make direct operational test, without disabling the remainder of the system.

In designing the New Jersey Turnpike communication system, many problems were encountered over and above the facilities which we have described, but to attempt to cover them all would be a large task. However, we believe that we have described the more important features, which adequately display the typical problems encountered in large-system development.

No Discussion

A Million-Watt Naval Communication Transmitter

J. C. WALTER
MEMBER AIEE

Synopsis: This article describes one of the world's most powerful radio telegraph transmitters. Recently built for the United States Navy, it is conservatively rated at 1,000-kw output over the 15- to 35-kc frequency range. Its use will ensure reliable communication with the United States Fleet even when high-frequency services are interrupted by magnetic disturbances. Unique features include the first use of super-power electron tubes having a power gain of at least 250; 10-microsecond fault protection; a variable master oscillator with 0.001 percent frequency stability, and provisions for frequency-shift teletype signalling.

NAVAL communication service requires positive direct transmission to mobile forces at sea and aloft, wherever they may be. The United States Navy is building a super-power radio transmitting station in the Pacific Northwest. It will be fitted out with the very-low-frequency (VLF) transmitting equipment to be described. Signals from this station will provide world-wide radio coverage in the lower half of the 15- to 35-kc band. Utmost reliability and continuity of service is ensured through careful planning and duplication of essential major components of the equipment.

Very efficient and economical high-frequency (HF) circuits are capable of handling the bulk of military traffic most of the time. They are, however, occasionally interrupted by magnetic and ionospheric disturbances. HF signals are severely attenuated when their circuit paths traverse relatively large auroral zones surrounding the earth's magnetic poles. VLF transmission over great distances requires the use of very high power to overcome attenuation and inherently poor antenna efficiencies. Its prime virtue is its ability to get through when higher frequencies are blacked out. VLF transmission has proved extremely reliable in the high latitudes.

Equipment

The basic transmitter equipment is known as the model *AN/FRT-3* radio transmitting set. It consists of the following principal apparatus:

1. Transmitter units in compartmented shielding;
2. Unified front enclosure;
3. Supervisory console;

Paper 53-199, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted December 26, 1951; made available for printing March 30, 1953.

J. C. Walter is with the Radio Corporation of America, Camden, N. J.

Fig. 1. Transmitter in test position at the factory showing supervisory console, and port and starboard frequency generator-monitors

4. Frequency generator-monitor;
5. Centralized control and distribution unit;
6. Metal-clad switchgear;
7. A-c load-center apparatus;
8. Cooling system;
9. Water-cooled test load;
10. Antenna tuning equipment

An efficient catenary antenna system was specially designed for suspension between two mountain ridges. A radial system of buried ground conductors contains nearly 200 miles of copper wire. Design of the antenna/ground system and materials for its construction were furnished as part of the equipment contract.

Description of Equipment

A general view of the transmitter in its test position at the factory is shown in Fig. 1. Identical port and starboard units are symmetrically disposed about a central control section. Port and starboard groups each consist of the following:

1. Frequency generator-monitor;
2. Main rectifier;
3. Exciter;
4. Power amplifier.

Port and starboard elements of the supervisory console and the control and distribution unit are consolidated in common assemblies.

For full-power operation both power amplifiers are driven by either exciter, and either frequency generator-monitor may be used to furnish keyed signals to the exciter in use. Main rectifiers have sufficient capacity for either unit to supply plate power to both power amplifiers. Idle units are thus in a ready reserve status and available for maintenance and testing.

Either power amplifier may be operated with either exciter to provide an output of 500 kw. The idle amplifier may be simultaneously operated with its test load up to 500 kw.

The supervisory console, Fig. 2, contains extension controls and instruments for trimming amplifier and antenna circuits. Selector switches in a mimic bus control various cross-switching operations. Mimic bus lamps show which transmitter units are in service and which are in idle or test status.

Port and starboard frequency generator-monitor (FGM) units are grouped with the supervisory console to provide a central operating position from which important instruments may be observed. The function of the FGM is to generate stable radio frequency voltage at a level sufficient to drive the transmitter to full power. Any frequency in the VLF range may be set up on the frequency generator. When the monitor is set to the same frequency, the generated frequency may be continuously monitored and automatically controlled to maintain a frequency stability of better than 0.001 percent. The FGM assembly includes a frequency-shift keyer with plus and minus deviation which is adjustable from zero to 25 cycles. The useful range of frequency shift is restricted by the extremely sharp resonance characteristic of the antenna. It is anticipated that at 25 cycles off resonance the radiated signal will be from three to six decibels below the level at resonance. Provisions are included in the FGM for either on-off or polar keying. The keyed r-f output of the FGM is fed through coaxial cable to the input terminals of the exciters.

Central enclosure master panel instruments include flow, pressure, and temperature indicators for the hydraulic and air systems. Master control switches for duplicate cooling systems are centrally located on this panel. Selector switches control the main rectifier ratio tap adjusters. Nine equal tap positions between 40 per cent and full voltage are registered on selsyn indicators. The ratio taps may be changed under load so that output power is adjustable at will without interruption of service.

Enclosure sections associated with port and starboard transmitter units include observation windows and interlocked access doors. Essential instruments and local controls are appropriately grouped on the various enclosure sections. Safety features include the following: door latch operated switches which trip rectifier circuit breakers and release gravity operated d-c ground devices; barrier door switches which additionally ground d-c buses; and insulated grounding sticks.

Compartmentation and radio-frequency shielding for rectifiers, exciters, and power amplifiers are provided by aluminum panels attached to structural aluminum framing behind the front enclosure. Aluminum interconnection wiring ducts and catwalks are mounted on top of the shielding structure just behind the upper section of the front enclosure.

Fig. 3 shows one of the high-power rectifier units which furnishes plate power for either or both power amplifiers at 11.5 kv. The rectifier is housed in a 10 by 20 foot compartment of the aluminum shielding structure. Seven type-870A

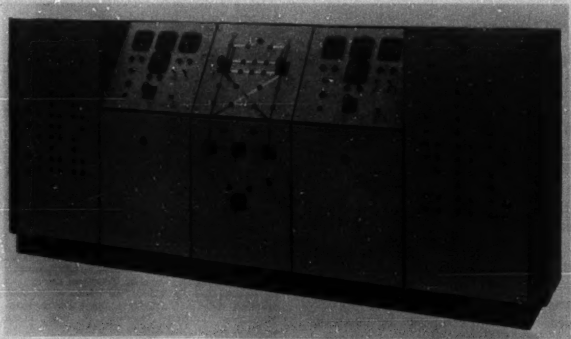


Fig. 2. Supervisory console

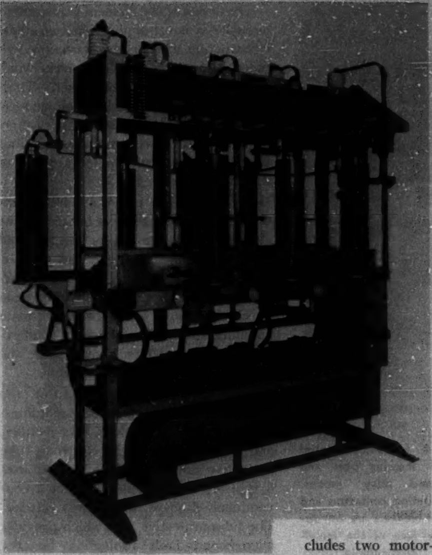


Fig. 3. Main rectifier tube assembly. (Note spare tube mounted on left end)

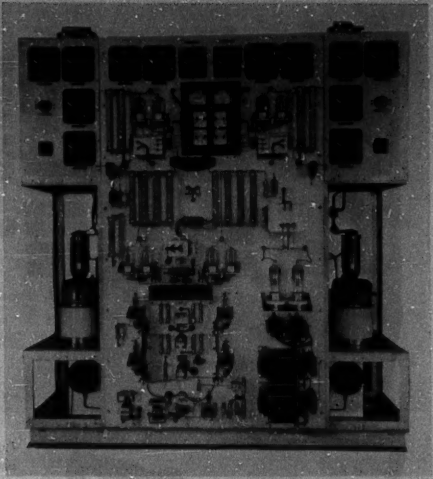


Fig. 4. Untuned radio-frequency exciter

metal-clad hot cathode tubes are mounted with their auxiliary apparatus and air heating and manifolding devices. Six of the tubes are normally connected in a conventional 6-phase double-way rectifier circuit.[1] The seventh is a preheated spare which may be substituted for any active tube by means of sparing switches mounted on a distribution bus. Cooling air is obtained from a station air-duct system located below the deck.

One of the exciters is shown in Fig. 4. This is a broad-band amplifier requiring no tuning over the entire VLF band. The output stage is coupled to either or both power amplifiers through an iron-core oil-cooled transformer. The unit is located within a 10- by 20-foot shielding compartment together with its plate supply rectifier and test load. The exciter rectifier unit consists of 6 type-*673* hot-cathode tubes in a conventional 6-phase double-way circuit and 4 type-*8008* hot-cathode tubes in two conventional diametric-rectifier circuits. The test load consists of a bank of wire-wound resistors mounted in a cabinet. Cooling air for the type-*892R* output tubes, the rectifier, and the test load is obtained from the station air system.

A general view of the port power amplifier radio-frequency circuit elements is shown in Fig. 5. The coil assembly includes two motor-operated variometers and manually adjustable output coupling inductors. The windings are low-loss Litzendraht conductor consisting of 16,128 strands of 5-mil enameled copper cabled around a jute core. The diameter over the glass braid is approximately 2 inches. Porcelain shafts mounting the variometer rotors are turned by motors located in shielded pits below the assembly. Stator windings are lashed to porcelain tubes supported in aluminum castings. Aluminum supporting columns are grounded at the bottom and bolted to bus-support insulators at the top to minimize coupled-in losses. All special porcelains and the Litzendraht conductor used in these assemblies are interchange-

Fig. 5. Power amplifier tubes and radio-frequency circuit elements. Output coupling coils are in center of plate variometer assembly

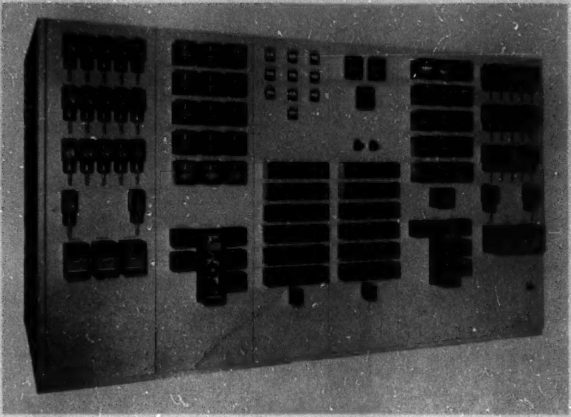


Fig. 6 (above). Control panel showing port and starboard relay groups. Distribution contactors and circuit breakers are on opposite face of the metal-clad cubicle

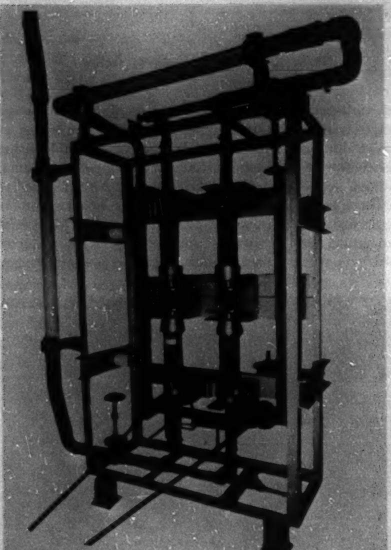


Fig. 7 (left). Water-cooled test load which dissipates 600 kw at radio frequencies

able with identical parts used in older 500-kw naval installations.

The conventional push-pull plate-tank circuit is tuned by push-button control of the plate variometers. Selection of 12 discreet frequency ranges over the 14.5 to 35 kc band is provided by air disconnects mounted on a tank capacitor bus. Forty oil-mica fixed capacitors are used in each power amplifier.

The tube frame assembly supports porcelain water coils, cooling water manifolds, and elements of the grid-tank circuit. Two of the three type-*5831* superpower tubes supported in their toroidal filament transformers are connected in push-pull. The third tube is an installed spare which may be quickly switched into service in lieu of either active tube. Motor-driven grid variometers are remotely controlled for push-button tuning of the grid circuit.

The design of the filament transformers embodies a novel application of toroidal construction wherein the concentric cylinders comprising the high current winding are used as the tube socket. These 14-kw units have a nominal voltage ratio of 175-to-6, and are used with reactors in series with their primaries to limit tube-starting currents to 150 per cent of their final value.

The type-*5831* tubes require exceptionally low driving power because of the electron optical principles incorporated in their design. Their use makes possible a considerable saving in both transmitter construction and operating costs by elimination of a 100-kw penultimate amplifier stage. An additional load reduction of over 50 kw is realized by the use of thoriated tungsten filaments.

Other assemblies located within the power amplifier compartment are a filter-capacitor rack, a shielded bias rectifier, and a cabinet housing an electronic-tube protective unit. The bias rectifier uses 6 type-*673* hot-cathode tubes in a conventional 6-phase double-way circuit. The tube protective unit which provides microsecond fault protection for the power amplifier tubes is a new development which will be described in the next section.

A metal-clad control and distribution unit is located behind the enclosure master panel. Control relays are centralized in two groups associated with port and starboard transmitter elements. Distribution contactors and branch-feeder circuit breakers are similarly grouped on the opposite side. A front elevation of the control panel is shown in Fig. 6. Main rectifier feeder overcurrent relays are drawout inverse-time units with instantaneous attachment. Selective automatic reclosure is provided by motor-operated reclose relays. Plunger-type overcurrent relays protect subordinate feeders.

The various equipments thus far described constitute the basic transmitter. When installed as a unit they occupy a floor space 80 feet wide by 50 feet deep. Auxiliary equipment designed for installation at a lower level occupies approximately the same area directly below the transmitter. Functionally important auxiliaries are briefly described in the following.

The a-c load center apparatus consists of air-cooled indoor distribution transformers, phasing transformers and station-type induction regulators. An overhead 460-volt bus from the customer's source is used for primary distribution. Regulated and unregulated 230-volt and

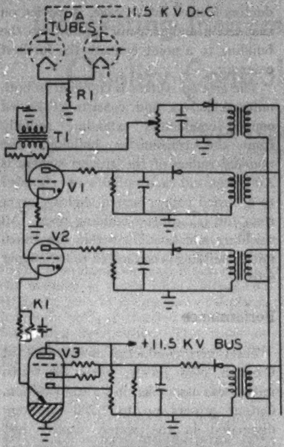


Fig. 8. Tube protective unit fundamental circuits

Table II. Typical Operating Data Recorded During 8-Hour Locked-Key Test

Function	Exciter, Port	Power Amplifiers		
		Port	Starboard	Combined
FGM Frequency, Kc per Second	14.80			
Exciter Plate Voltage, D-C Kv	10.05			
Plate current no. 1, d-c amperes	0.36			
Plate current no. 2, d-c amperes	0.34			
Plate power input, kw	7.00			
Power Amplifier Plate Voltage, D-C Kv		11.50	11.50	
Total plate current, d-c amperes		61.00	64.00	
Plate power input, kw		700	735	1,435
Bias no. 1, d-c volts		600	780	
Bias no. 2, d-c volts		600	650	
Bias resistance no. 1, ohms		700	550	
Bias resistance no. 2, ohms		800	500	
Grid current no. 1, d-c amper.es		0.22	0.06	
Grid current no. 2, d-c amperes		0.08	0.40	
Plate tank current, r-f amperes		260	262	
Test load resistance, ohms		8.5	8.5	
Power output, kw		572	583	1,155
Plate efficiency, per cent		81.7	79.2	80.5
460-Volt Feeder Load, Kw				110
Power factor, lagging				0.87
4,160-Volt Feeder Load, Kw				1,550
Power factor, lagging				0.94
Total Demand, Kw				1,660
Key up load, kw				102
Standby load, kw				95

460-volt secondary cables run to the distribution unit above and to other loads. Starting reactors for high-power filament circuits are mounted in a frame with associated power-factor correction capacitors.

Certain high-voltage d-c apparatus is also located adjacent to the a-c load center. It consists of port and starboard rectifier filter reactors, current-limiting and transient-suppressor resistors, and a motor-operated d-c transfer switch. The latter is used for cross-switching of the main rectifiers. An expanded metal protective fence encloses the area occupied by the a-c load center and high-voltage d-c apparatus.

Port and starboard blowers furnish air to a station air duct previously mentioned. Either blower may be used alone for this purpose. Each is enclosed in a separate room provided with metal intake filters. Approximately 8,000 cubic feet per minute air delivery, at 1.25 inches static pressure, is required for cooling purposes.

Table I. Test Performance Versus Salient Specifications

Performance	Specification	Test
Locked key demand, kw	1,770	1,660
Average load, 20 words per minute keying, kw	1,105	1,037
Standby load, kw	105	95
Rectifier d-c regulation, per cent	10	9.85
Filter output ripple, rms, per cent	0.50	0.41
Output radio-frequency power, 12 frequencies, kw	1140	1,150
Keying speed, dot-cycles per second	75	75
Frequency stability, per cent	0.001	0.0003

Water-cooled test loads for the port and starboard power amplifiers are located in the blower compartments. Each consists of four water cooled elements mounted in the assembly shown in Fig. 7. Each load is rated at 600 kw dissipation with a flow of 180 gallons per minute of raw water coolant.

Pure water used for cooling the power amplifier tubes is stored in two 1,000-gallon copper tanks. A closed circulating system pumps approximately 120 gallons per minute to each power amplifier. Heated water returns by gravity to the storage tanks. Raw water passing through heat exchangers cools the pure water. All pumps are installed in duplicate, and furnished with appropriate valving to provide for maintenance without interruption of service. As a battle-damage precaution against possible loss of the raw water supply a complete air/water heat-exchanger system is provided as an installed stand-by equipment.

Three-phase plate transformers for the main rectifiers are oil-filled units installed in a typical vaults. Each is rated at 1,700 kva, 5.5 per cent impedance, and 4,160 a-c phase voltage. Small plate transformers for the exciter rectifiers are installed in the same vault.

A switchgear cubicle consisting of a metering section and two high-speed air circuit breakers is installed adjacent to the vault. The circuit breakers are in the 4,160-volt feeders to the main rectifier transformers. They are draw-out units rated at 600 amperes, 5,000 volts, with an interrupting capacity of 100,000 kva. The switch-gear unit is designed for installation in line with matching customer's equipment.

Tube Protective Unit

The tube protective unit is a high-speed electronic control device used to protect the power amplifier tubes from damage due to internal gas arcs or similar faults. Whenever there is a sudden rise in power-amplifier tube current above normal, the d-c plate-supply bus is shorted to ground through a type-*5630* ignitron. This protection is applied 7 to 10 microseconds after the fault is initiated.

Fig. 8 shows the fundamental circuits of this unit and its connection to the power amplifier. Under normal operation, the unit is in the ready condition. Should a tube fault occur, a sudden rise in current above normal causes a rise in voltage to appear across the power-amplifier common cathode resistor *R1*. This is applied through pulse transformer *T1* to the grid of thyratron *V1*. When *V1* fires, it, in turn, fires a larger thyratron *V2*. The current through *V2* is of sufficient magnitude and duration to cause ignition of *V3*. Since the main anode of ignitron *V3* is permanently connected to the d-c bus, the power-amplifier tube fault current that initiated this sequence is transferred to the ignitron. At the same time, relay *K1* is energized and trips the circuit breaker in the main rectifier feeder. Backup protection is provided by primary a-c overcurrent relays, whose contacts operate in the same trip circuit. The feeder circuit breaker is opened in approximately 3 cycles on a 60-cycle basis after its trip coil is energized.

The time interval from fault occurrence to the application of protection was determined by electronic methods, and consistently was found to be 7 to 10 microseconds. Transient calculations indicate that this provides a generous margin of safety in preventing the build up of currents that could damage the power-amplifier tubes.

Antenna Tuning Equipment

Port and starboard antenna equipments each consist of a motor-driven variometer, two helix assemblies, an antenna lead-in trunk, and a power-operated grounding switch. The variometer is similar to, but somewhat larger than, one of the power amplifier variometers. Its windings consist of two parallel 2-inch Litzendraht cables supported on porcelain tubes mounted in shielded concrete piers. A tuning motor and reduction gear are installed in a shielded pit below the vertical rotor shaft. The series-loading helices are wound with two parallel Litzendraht cables, and sectionalized to provide appropriate tap points for major changes in the operating frequency. Helix assemblies are approximately 14 feet in diameter by 26 feet high. The winding are assembled on stacked porcelain spacers. An antenna lead-in trunk consists of a fabricated copper spar 82 feet long by 28 inches in diameter. The trunk is supported by suspension insulators above the helix area and extends outdoors through a wall port 20 feet square. The spar is articulated to take care of slight movements of the antenna feeder bus downlead. A rat-tail cage is brought down to the lead-in trunk from the antenna bus which is dead-ended on a nearby tower. A fabricated grounding-switch arm lowers from the overhead shielding to make contact with a pad on the lead-in trunk.

Port and starboard helix rooms are 75 feet square by approximately 60 feet high, and completely copper-lined. Such dimensions are necessary to minimize losses. Currents of approximately 1,000 amperes in each antenna half are anticipated at the lower operating frequencies. Voltage at the top of the helices is expected to approach 200 kv. Low-loss shielding of appropriate thickness is necessary to prevent extreme losses in reinforcing steel. Structural damage to insufficiently shielded buildings housing similar tuning equipment has recently been reported.[2]

Antenna/Ground System

The design of the antenna/ground system was based upon topographic limitations peculiar to the installation site. Each half of the antenna consists of five catenary suspensions supported from 200-foot steel towers erected on opposing mountain ridges. The arrangement provides five separate flat-top capacity elements having an average elevation of approximately 1,000 feet over the valley floor. A downlead from each flat-top section is connected to a common radio-frequency feeder bus connecting the transmitter output to the radiating system. Average spacing between individual catenaries in the port and starboard groups is roughly 400 feet, while the inboard spans of the two groups are separated by about twice this distance. The transmitter building is located between the antenna halves at a point of low elevation. Each catenary suspension consists of an active radio-frequency portion made up of approximately 4,500 feet of 1-inch-diameter stranded copperweld conductor socketed on each end, and attached to appropriate insulator assemblies. Suspension tail cables of 1-inch-diameter galvanized steel strand are dead-ended on the 200-foot towers on opposite sides of the valley. Average length of the catenaries is 7,700 feet; maximum length, 8,800 feet. Nominal working strain is 31,000 pounds stringing tension, and 44,000 pounds fully loaded with 1/2-inch radial ice and a wind pressure of 8 pounds per square foot of projected area. Rated breaking strength is 93,000 pounds. Strain insulator assemblies consist of two quadruple yoke assemblies in tandem at each end of the radio-frequency portion of each span. A quadruple yoke assembly consists of four porcelain tube strain-insulator units in parallel, each 72 inches long by 6 inches diameter. Each end of a tandem group is provided with aluminum corona rings. Tail cables are fitted with conventional vibration dampers near the dead-end attachment point. A downlead taken off near the center of the copperweld section consists of two portions. The upper is lightweight hollow-copper conductor, approximately 1 inch in diameter and averaging 1,000 feet in length. The lower section of 400 feet consists of two parallel conductors spaced to minimize corona loss. A radio-frequency bus, consisting of a group of conductors formed into a cage, is carried on suspension and tangent towers from the building to a point beneath the furthest catenary.

The ground system is common to both antenna halves and consists of buried copper conductors extending radially from the transmitter building. The nominal radius of the ground system is 3,000 feet and each radial wire is bonded to a buried copper screen slightly larger than the transmitter-building area. All shielding and bonding within the transmitter building is connected to the copper screen.

Performance

Tests were conducted during 1950 at Camden, New Jersey, where the transmitter was assembled in its test position. Output powers up to 1,200 kw were dissipated in the water cooled test loads. Test data covering performance against salient specifications are listed in Table I. Typical data recorded during an 8-hour locked-key test are given in Table II.

Final phases of circuit development included a search for parasitic impedances. A wide-range oscillator and cathode-ray oscilloscope were used to seek out grid- and plate-circuit resonances, and to determine whether their reactances were of the proper sign to support oscillations at spurious frequencies. One extremely-low-frequency resonance in the grid circuit was eliminated by substituting series bias feed for the original shunt feed. Most of the parasitic circuits found at frequencies above the operating range were of the shock-excited variety. This was deduced from observations of damped-wave oscillations appearing in the cathode current pulse during dynamic operation. A final arrangement of series-damping resistors in the grid circuits, and a resistor-capacitor series suppressor from each anode to ground, effectively damp these spurious circuits. Full-power keying tests, at speeds up to 160 dot-cycles per second (400 words per minute), revealed no signs of instability.

References

1. Pool Cathode Mercury-Arc Power Converters. *Publication No. C34.1—1949*, American Standards Association, New York, N. Y., 1949, pp. 11–14.

2. The Giant Station at Jim Creek, J. R. Redman. *Signal*, Army Signal Association, Washington, D. C., Jan.-Feb. 1951, pp. 15–60.

◆

No Discussion

Integrating Instruments for Simplified Quality-Control Measurements

D. M. LONGENECKER
MEMBER AIEE

Synopsis: In quality control work, it is often necessary to determine the mean and the standard deviation of an important variable. To accomplish this, a measurement system has been developed for applications in which the material is tested while moving continuously, most of the necessary calculating functions being performed by two new integrating instruments. This paper describes the application of such a system to the measurement of yarn thickness.

SINCE VARIATION is virtually inevitable in most manufacturing processes, quality control necessarily includes the three steps of measuring, studying the variations, and controlling the variations. This paper is concerned with the first two steps, which are usually accomplished by making numerous individual measurements of the important parameter, followed by some sort of statistical study of the data. However, this ordinary procedure has three serious disadvantages.

1. It may require appreciable time and expense to make the numerous measurements and to record and compute the results;
2. The computation and interpretation of the data is often not well understood by manufacturing personnel having little or no knowledge of the mathematics of statistics;
3. There is an uncertainty in the results whenever small-lot sampling methods are used, because of normally expected chance differences between the characteristics of a small sample and those of the larger production lot from which the sample was taken.

This third matter depends upon the size of the sample, and it involves the study of sampling distributions, which will be found in most textbooks on statistics.[1]

Undoubtedly there would be a wider application of scientific quality control procedures, if the results of the statistical analysis could be indicated immediately, and if larger samples could be tested economically so there would be a high degree of certainty that the production lot is not significantly different from the sample. To achieve these objectives, a system has been developed to perform an integrated measurement of a dimension of a continuously-moving sample. This measurement system, which will be described in this paper, has been named an Evenness Integrator, and is shown in Fig. 1.

Mean and Standard Deviation

Shewhart[2] has shown that the best simple criteria of quality for a varying parameter are the mean $\overline{X}$, and the standard deviation σ. The purpose of the Evenness Integrator is to provide a quick and convenient means for determining these two important statistics of the sample. To illustrate, consider the testing of yarn and similar materials in the textile industry. The mean thickness of yarn must conform, within certain limits, to the established standards for that particular material. Furthermore, the degree of variation (unevenness) of yarn thickness is another important criterion of quality, and the standard deviation σ is a good measure of this characteristic.

One testing device,[3] which is widely used in textile mills in this country, is known as the Pacific Evenness Tester, and it provides a chart record of the instantaneous thickness of a moving sample of yarn. The yarn passes between two rolls at a substantially constant speed of 12 yards per minute, as shown in the left part of Fig. 2. Since the lower roll is grooved and the upper one has a mating tongue, the material is confined in the lateral direction. Carried on the hinged and weighted arm is the upper roll, which is free to move up or down in response to the varying thickness of material between the rolls. The primary detector is an electric dimension gauge of the variable-inductance type, responding to the continually varying thickness as the material passes between the rolls. A micrometer screw, which is provided on the electric gauge as a means of suppressing the zero, is ordinarily set in advance to the expected or "correct" thickness of the material, thickness deviations being measured from this preset value.

A chart record, such as Fig. 3, will show the general pattern of thickness variation and may provide much useful information, but it does not enable one to express the mean or the amount of the variation in numbers. Therefore, it is necessary to obtain the mean and the standard deviation of the thickness. To accomplish this, two new end devices are provided in the Evenness Integrator.

Integrating Instruments Used

The first of these, a d-c current-integrating instrument, indicates the mean thickness deviation from the preset value. The second, an a-c current-squared integrating instrument, indicates the mean square of the thickness deviation. As will be shown later in this paper, there are simple relations between the readings obtained from such integrating instruments and the desired statistics. They are as follows

mean $=\overline{X}=A+G$

and

standard deviation $=\sigma=\sqrt{B-A^2}$

where

A = the reading of the current-integrating instrument
B = the reading of the current-squared integrating instrument
G = the setting of the gauge micrometer

The foregoing calculation of the mean can usually be done mentally. A nomograph is provided for calculating the standard deviation, so this conversion also becomes a simple matter for the operator. Thus, the integrating technique eliminates virtually all of the laborious data-taking and the computations ordinarily associated with the statistical analysis of a large amount of data obtained from the test of a large sample.

Functional Principles of System

The functional principles of the integrating system are shown schematically in Fig. 2. The windings of the electric dimension gauge are connected to form two arms of an a-c impedance bridge, the zero of this bridge being set for zero voltage output when the micrometer screw is on zero and the rolls are empty. The micrometer is then reset to the approximate thickness expected of the yarn, and the yarn threaded into the rolls and brought up to speed. Any upward mo-

---

Paper 53-205, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15-19, 1953. Manuscript submitted March 17, 1953; made available for printing April 2, 1953.

D. M. LONGENECKER is with the General Electric Company, Lynn, Mass.

The author gratefully acknowledges the co-operation of the Engineering Research Department of Pacific Mills, Lawrence, Mass., and the Anderson Machine Shop, Inc., Needham Heights, Mass.

tion of the steel armature has the effect of increasing the inductance of the upper coil while decreasing that of the lower one. This produces a 60-cycle alternating voltage e proportional to the amount of deflection of the armature from its midposition. A downward deflection produces a similar voltage output, but in this case the phase of the voltage differs by 180 degrees.

The voltage e, which is about 10 millivolts per thousandth of an inch of gauge deflection, is amplified by a conventional electronic amplifier to provide an a-c output i, which varies in phase and magnitude according to the thickness deviation from the micrometer setting. The output current i is approximately 7 milliamperes for 0.001-inch gauge deflection, when the attenuator is set to the highest-gain position. Taking the path indicated by heavy lines in Fig. 3, this current flows through the squaring integrator, which moves always in a clockwise direction and at a speed proportional to the square of the current.

Simultaneously, an alternating voltage V, proportional to i, drives a phase-sensitive a-c to d-c converter, which employs germanium rectifiers. It produces a d-c output proportional to V in magnitude, and having a polarity governed by the phase of V. This d-c output I_{DC}, is of the order of 0.7 milliampere for 0.001-inch deviation, assuming the attenuator is set to the highest-gain position. It flows through the current-integrating instrument and an external recorder, as indicated in Fig. 3. The current integrator has a zero-center scale, and moves clockwise or counterclockwise, depending on the polarity of I_{DC}, and at a speed proportional to the current.

An important part of the system is a timer, which controls the duration of the integrating cycle. After a preset time interval, the timer opens contact 1 and closes 2, de-energizing the relays and opening all four of the relay contacts (shown closed in Fig. 3), thereby disconnecting both integrators from the circuit. The pointers of these integrators then stop immediately, and their readings are recorded by the operator.

Closing of the timer contact 2 energizes the reset circuit, which is essential for resetting the instruments to zero prior to starting a new test. "Right" and "left" reset controls are provided to control the d-c reset voltage PQ for the current integrator, which has a zero-center scale. Operation of either of these controls also applies the a-c reset voltage RS to the squaring instrument, producing a counterclockwise rotation of its pointer, which resets against a pin at the zero end of the scale. Resetting can be completed in twenty seconds or less.

To initiate a new integrating cycle, a start button (not shown) is pushed, closing timer contact 1 and energizing the two relays, so that their four contacts again connect the integrators to the circuit, as shown in Fig. 3.

Continuous or timed operation can be selected for the recorder by closing or opening the switch SW, since dummy load resistors N and O are switched into the circuit by the relays, replacing the integrators while the latter are disconnected. The integrating time, which can be preset to any desired value from 15 to 150 seconds, is controlled by the adjustable timer. While the equipment is ordinarily calibrated for a standard interval such as 2 minutes, other time intervals can be used if the instrument readings are multiplied by appropriate factors. A gain control in the amplifier provides the calibrating adjustment for the system.

It should be noted that the two integrating instruments, the end devices of this measurement system, operate differently from the familiar indicating instruments. Having practically no spring torque, their pointers will not move appreciably so long as no current flows, regardless of the position of the pointer. The pointers are set to zero prior to starting the measuring cycle, and the speed of pointer motion is governed by the current. It is the final position of the pointer, at the end of the test, which is read and used.

Continuously Moving Samples Tested

The basic principle of testing continuously-moving samples, while not new, has certain particular advantages in quality control measurements. Since materials can be tested at fairly high speed with this method, relatively large samples can be tested economically. Furthermore, the system responds continuously to the varying thickness, so that mathematically the effect is as though an almost infinite number of thickness measurements were made on the sample. These factors ordinarily make it safe to assume there is no appreciable difference between test results on a single sample and the production lot in question, provided that the sample is unbiased.

Performance

Sensitivity of the equipment used in the textile industry is given in Table I for the most commonly used integrating interval, 2 minutes. In this application, the scale of the current integrator is marked -2 to 0 to $+2$ in thousandths of an inch and the squaring integrator has a scale of 0 to 7, making the equipment direct-reading for the times-1 attenuator setting. The system is used for materials varying from 0.004 to 0.06 inch in mean thickness, with standard deviations ranging from 0.001 to 0.012 inch.

The measuring system has been found to respond satisfactorily to instantaneous values of $X-G$ ranging from 0.05 times to over three times the full scale values listed in column 2 of Table I. This 60-to-1 range provides adequate accuracy of integration for normal random-varying signals, so long as a favorable degree of amplification is used by setting the attenuator properly. Accuracy of ±6 per cent of full scale is achieved, providing

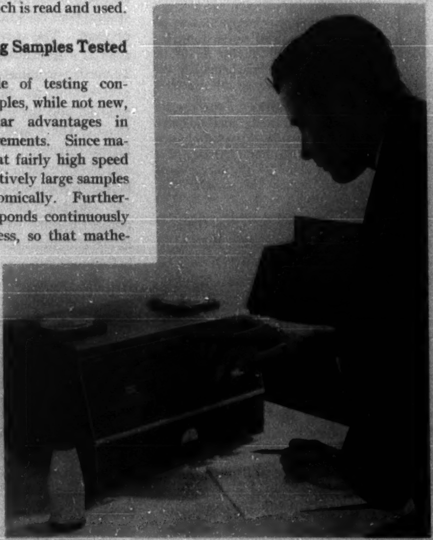


Fig. 1. Evenness Integrator

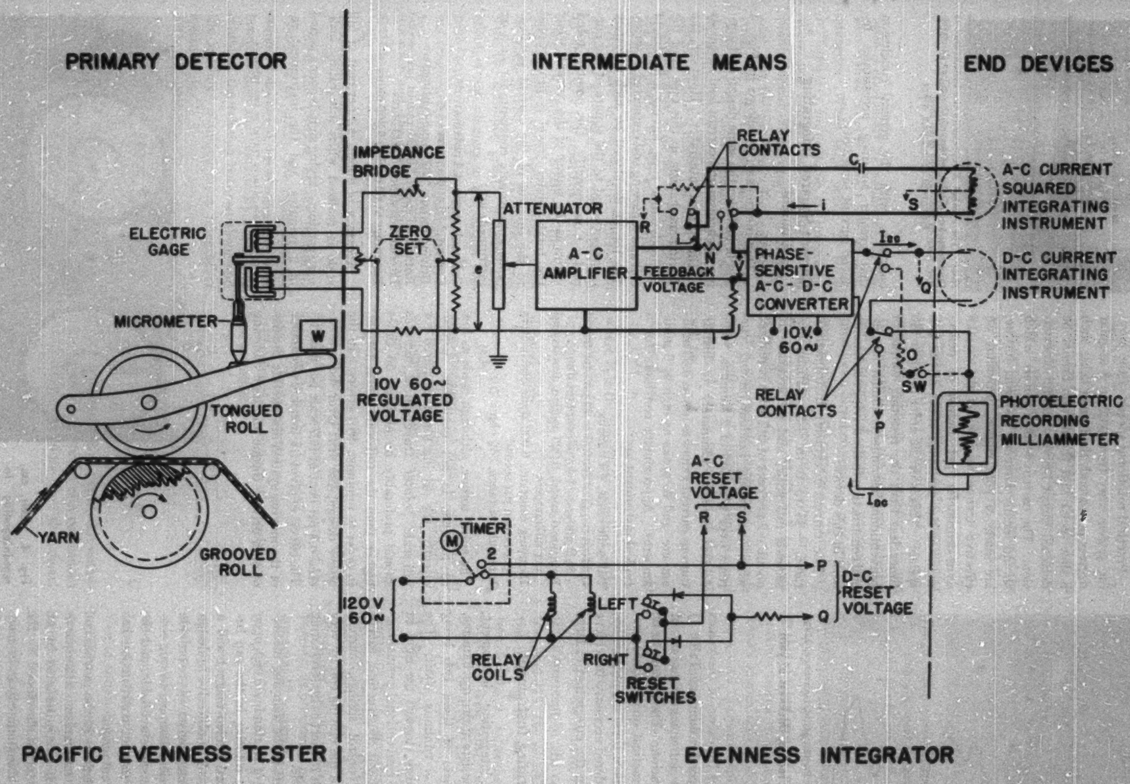
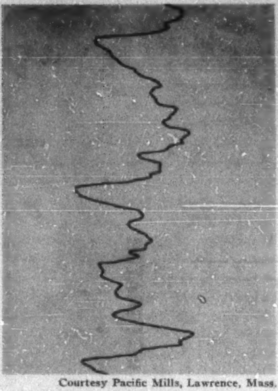

PRIMARY DETECTOR
INTERMEDIATE MEANS
END DEVICES
ELECTRIC GAGE
MICROMETER
W
TONGUED ROLL
GROOVED ROLL
YARN
IMPEDANCE BRIDGE
ZERO SET
10V 60~ REGULATED VOLTAGE
e
ATTENUATOR
A-C AMPLIFIER
R
FEEDBACK VOLTAGE
RELAY CONTACTS
N
V
C₁
i
S
PHASE-SENSITIVE A-C - D-C CONVERTER
10V 60~
Idc
Q
O
SW
P
RELAY CONTACTS
A-C CURRENT SQUARED INTEGRATING INSTRUMENT
D-C CURRENT INTEGRATING INSTRUMENT
PHOTOELECTRIC RECORDING MILLIAMMETER
Idc
TIMER
M
2
1
120V 60~
A-C RESET VOLTAGE
R
S
P
Q
D-C RESET VOLTAGE
LEFT
RIGHT
RELAY COILS
RESET SWITCHES
PACIFIC EVENNESS TESTER
EVENNESS INTEGRATOR

Fig. 2. Functional diagram showing the new Evenness Integrator

Details of the Pacific Evenness Tester are included by courtesy of Pacific Mills, Lawrence, Mass., and Anderson Machine Shop, Inc., Needham Heights, Mass.



Courtesy Pacific Mills, Lawrence, Mass.

Fig. 3. Chart record, thickness of yarn

high peak values (in excess of the values listed in the second column of Table I) occur relatively infrequently or are of fairly short duration. The signals to be integrated in practical quality measurements usually have these characteristics, so it is not considered necessary to achieve high accuracy in the integration of large peaks.

Current-Integrating Instrument

The current-integrating instrument, Figs. 4 and 5, consists of two mechanisms, the moving parts of these two being connected by a common shaft. The driving mechanism, toward the right in Fig. 5, is of the conventional permanent-magnet moving-coil construction, except there are no control springs, current being brought to the moving coil by very light spirals built to exert as little restoring torque as possible.

Toward the left is the damping mechanism, consisting of an aluminum cup, open at the left end, rotating in the gap between an internal Alnico magnet and a steel outer ring. The objective in design of the damping mechanism was to provide the largest practical damping torque consistent with the space available, and the 4-pole internal magnet construction has been found advantageous.

When current flows in the moving coil of the driving mechanism, a torque is developed which is counteracted by an equal damping torque developed by the rotation of the aluminum cup in its magnetic field. This damping torque is directly proportional to the rotational speed, while the driving mechanism develops a torque proportional to the current. The result is a speed of rotation proportional to the current.

Practical rotational speeds of the moving system vary from 0.02 to 8 angular degrees per second, the former being such a slow speed that the motion cannot be detected except by very careful observation. The current lead-in spirals, which should ideally have no restoring torque, are made of gold ribbon, providing an effective torque as low as 0.4 to 0.8 dyne-centimeter per radian of deflection, for the combined effect of the two spirals. This is of the order of 70 to 140 times less than the restoring torque of conventional instruments similar in general construction to the driving mechanism used in this device.

The current-integrating instrument is rated −85 to 0 to +85 milliampere-seconds full scale, and an accuracy of about ±2 per cent of full scale is realized when integrating random-varying currents over a period of 2 minutes. The device responds adequately to instantaneous current values as low as 0.03 milliampere, and to peaks as high as 6 milliamperes. Approximate constants for a typical instrument are:

Full-scale angle, 4.4 radians
Current required for full-scale pointer travel in 2 minutes, 1.4 milliamperes d-c
Rotational speed, 0.025 radian per second per milliampere
Damping constant, 6,000 dyne-centimeters per radian per second
Internal resistance, 340 ohms

The spiral torque, although made as low as possible, is not negligible. Consequently, the driving torque required for full-scale pointer travel in 2 minutes should preferably be at least 50 times higher than the full-scale spiral torque, in order to minimize drift errors. In the above design, the drift error did not exceed 0.5 per cent of full scale per minute. The effects of pivot friction and unbalance of the moving system also dictate the use of high driving torque and high damping constants, with the general conclusion that greater accuracy necessitates reduced current sensitivity.

Table I. Evenness Integrator Sensitivity, 2-Minute Interval

Attenuator Setting, M	Mean Thickness Deviation in Thousandths, for Full Reading of Current Integrator $\overline{X}-G$	Mean Squared Thickness Deviation in Thousandths, for Full Reading of Squaring Integrator $\overline{(X-G)^2}$
0.5	−1 to 0 to +1	1.75
1	−2 to 0 to +2	7
2	−4 to 0 to +4	28
5	−10 to 0 to +10	175

Note: The symbol *X* represents the instantaneous thickness, and *G* the setting of the gauge micrometer, both in thousandths of an inch.

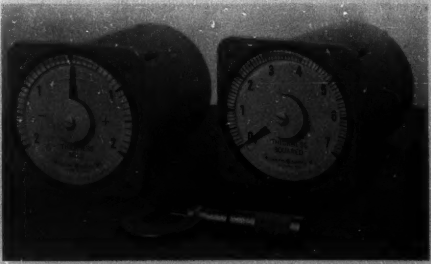
Current-Squared Integrating Instrument

This instrument, as shown in Figs. 4 and 6, includes the same design of damping mechanism and current lead-in spirals as the current integrator. However, the driving mechanism in this case is of the electrodynamic type, having a fixed-field coil and a moving coil connected in series. The torque developed by such a system is proportional to the square of the current. Consequently, the pointer speed varies as the square of the current, while the total pointer travel is proportional to the integral of the current squared. Some approximate constants for such an instrument are:

Full-scale angle, 4.4 radians
Current,* 10 milliamperes a-c
Driving torque,* 200 dyne-centimeters
Rotational speed,* 0.036 radian per second
Internal resistance, 1,600 ohms
Reactance at 60 cycles, 5,300 ohms

Accuracy of about ±3 per cent of full scale is realized when integrating ordinary random-varying signals over a period of 2 minutes, and the instrument re-

Fig. 4. Left, current-integrating instrument. Right, current-squared integrating instrument



---

* For full-scale pointer travel in two minutes.

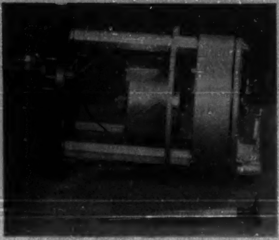


Fig. 5. D-c current-integrating instrument, scale and cover removed

sponds adequately to instantaneous current values ranging from 1.5 to 30 milliamperes. When large peaks, 20 to 30 milliamperes, for example, are integrated, there are instantaneous errors of the pointer speed of the order of 5 to 8 per cent, but the practical effect on the accuracy is slight, since such peaks are ordinarily of short duration, and they constitute only a small portion of the integrated current value. The drift error is about the same as for the current-integrating instrument, 0.5 per cent of full scale per minute.

As shown in Fig. 3, a 0.5-microfarad capacitor C is used with this instrument to provide a low-impedance load at 60 cycles, in order to minimize the volt-ampere burden upon the amplifier. Considerable inverse current feedback is used with the amplifier, greatly reducing the effect of any changes in the output current i occurring from changes in the winding temperatures of the squaring integrator. This feedback also improves the linearity and the gain stability of the system.

Mathematical Basis

Since the purpose of the measuring system is to obtain the mean and the standard deviation, these quantities have been examined mathematically, in order to establish the relation between them and the readings obtained from integrating instruments.

Shewhart[2] defines the mean as follows:

$$\bar{X}=\frac{X_1+X_2+\ldots+X_n}{n} \qquad (1)$$

where X_1, X_2, etc., are individual discreet measurements of the variable, and n the number of such measurements.

When the sample is tested while moving continuously, the signal e, Fig. 3, is a continuously varying voltage proportional to the instantaneous thickness deviation from the preset value. Consequently, n may be considered infinite. The definition may therefore be rewritten as

$$\text{mean}=\bar{X}=\frac{1}{t}\int_0^t X\,dt \qquad (2)$$

where t is the time duration of the test in minutes and X the instantaneous thickness of the material.

Referring to Fig. 3, the a-c output of the amplifier i can be expressed by the relation

$$i=k_1(X-G) \qquad (3)$$

where G is the setting of the micrometer in thousandths of an inch, and k_1 is a constant depending on the amplifier gain and the sensitivity of the primary detector and its associated circuit. The converter produces a direct current I_{DC}, proportional to the alternating current i. This direct current may be either positive or negative, depending on whether X is larger or smaller than G.

If a center-scale instrument responds to the integral of a direct current, its final reading A at the conclusion of the test will be

$$A=k_2\int_0^t i\,dt \qquad (4)$$

where k_2 is the sensitivity constant of the instrument and its converter. Substituting equation 3 into 4, and using the definition of the mean, equation 2, it can be derived that

$$\bar{X}=\frac{A}{tk_1k_2}+G \qquad (5)$$

If a standard test time is used for all samples, it is simple to calibrate the equipment so the product tk_1k_2 is one. Then

$$\bar{X}=A+G \qquad (6)$$

Thus, a d-c current-integrating instrument gives a convenient measure of the mean, if its reading is added to the setting of the micrometer.

Shewhart[2] also defines standard deviation as follows

$$\sigma=\sqrt{\frac{(X_1-\bar{X})^2+(X_2-\bar{X})^2+\ldots+(X_n-\bar{X})^2}{n}}$$

$$=\sqrt{\frac{X_1^2+X_2^2+\ldots+X_n^2}{n}-\bar{X}^2} \qquad (7)$$

For the purpose of this paper, this definition is rewritten in terms of a continuous variable, as was done for the mean. Thus

standard deviation =

$$\sigma=\sqrt{\frac{1}{t}\int_0^t X^2dt-\bar{X}^2} \qquad (8)$$

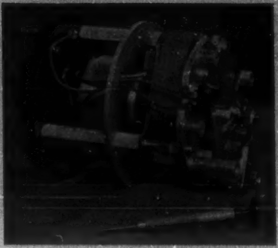


Fig. 6. A-c current-squared integrating instrument, scale and cover removed

It will be recognized that this is a difficult quantity to measure with instruments, since it involves not only an integral of X^2, but also $\bar{X}^2$, a quantity that is unknown prior to the test and which becomes known only upon completing the integrated measurement on the entire sample.

The mathematical expression for the response of a current-squared integrating instrument is the following

$$B=k_3\int_0^t i^2dt \qquad (9)$$

where B is the reading from the scale of the instrument at the conclusion of the test, and k_3 is the sensitivity constant of the instrument. By substituting equation 3 into 9, expanding the resultant expression, and using equations 2, 8, and 5, it can be proved that

$$\sigma=\sqrt{\frac{B}{tk_1^2k_3}-\left(\frac{A}{tk_1k_2}\right)^2} \qquad (10)$$

Constant k_1 is determined by adjusting the gain control so $tk_1^2k_3=1$. A second requirement is to adjust the converter sensitivity, which affects k_2, so that $tk_1k_2=1$, as previously mentioned. Then equation 10 becomes

$$\sigma=\sqrt{B-A^2} \qquad (11)$$

Thus, while it is not possible to indicate standard deviation with the comparatively simple instruments described here, it is possible to obtain the standard deviation from both of the instrument readings, A and B, by performing the calculation of equation 11.

When a variety of materials are to be tested with one equipment it is desirable to provide an attenuator switch, as in Fig. 3, this switch being designed for simple scale-multiplying factors such as 0.5, 1, 2, and 5. The multiplying factor can be designated as M, and it will vary according to the reciprocal of the change in gain effected by the attenuator. For such an arrangement, the modified form of equations 6 and 11 is

$$\bar{X} = MA + G \tag{12}$$

and

$$\sigma = M\sqrt{B-A^2} \tag{13}$$

When testing a given type of material, doubling the amplification doubles the reading of the current-integrating instrument, while quadrupling that of the current-squared integrator. The most favorable accuracy is obtained by the use of as much amplification as possible, so long as neither instrument reaches the end of its scale before the end of the time interval.

Conclusions

It is believed that integrating instruments will be found useful in the solution of various specialized measuring problems, of which determining mean and standard deviation is only one example. Obviously, it should be possible to construct instruments for integrating various different quantities by substituting various driving mechanisms for the two described, still using a moving system having essentially no restoring torque and a very high degree of damping. It will be recognized that an integrating instrument is actually a form of computing mechanism, and that the measuring system described herein is a type of analogue computer.

Two new end devices, a current-integrating instrument and a current-squared integrating instrument, have been incorporated, together with conventional parts including an electric dimension gauge, impedance bridge, amplifier, timer, relays, and converter, into a measuring system applicable to dimensional quality measurements for applications where the material is tested while moving continuously. The instruments indicate the mean and the mean-square deviation from a preset value, and mathematical derivations have shown that these readings can be easily converted to the mean and standard deviation of the measured variable.

Although the measurement system described here was developed for testing yarn, similar techniques could presumably be applied to other materials, such as paper, metal strip, wire or plastics, which can be tested while moving at a constant rate of speed. The measured variable could be thickness, width, diameter or any other electrical or nonelectrical quantity, provided that a suitable primary detector is available. It is believed that such instrumentation will reduce the time and effort required to measure certain qualities of manufactured products, encouraging a wider application of scientific quality-control procedures in industry.

References

1. Introduction to Statistical Analysis (book), **W. J. Dixon, F. J. Massey, Jr.** McGraw-Hill Book Company, Inc., New York, N. Y., 1951.

2. Economic Control of Quality of Manufactured Product (book), **W. A. Shewhart.** D. Van Nostrand Company, Inc., New York, N. Y., 1931.

3. Yarn Tester Uses Electronic Elements. *Electrical Manufacturing*, New York, N. Y., vol. 51, no. 2, Feb. 1953, pp. 142–44.

No Discussion

An Application of Boolean Algebra to the Design of Electronic Switching Circuits

S. H. WASHBURN
ASSOCIATE MEMBER AIEE

IN THE 15 years since Shannon[1] introduced the Boolean algebra as an aid towards the analysis and synthesis of switching circuits, this algebra has proved to be a worthwhile addition to the procedures available to the switching circuit designer. Although, in its present form, it cannot be employed to solve many of the problems by which the designer is confronted, the switching algebra provides an exceedingly effective technique in those cases where its use is indicated.

The utility of the Boolean algebra is evidenced in two ways. First, the algebraic symbolism is a useful notation for representing the behavior of a switching circuit resulting from various input conditions and its effect on the condition of the output leads which it controls. Second, the theorems of the algebra indicate possible manipulations of circuit configurations, thus allowing the designer to obtain various equivalent circuit forms without fear of the consequences of lapses of logic.

In this paper, a method utilizing the Boolean algebra in the design of electronic switching circuits is presented. This method is analogous to that applicable to the design of relay switching circuits in that it deals with the interconnection of idealized 2-valued building blocks, such as gates and memory units, into so-called logical circuit arrangements. No attempt is made to provide design tools for the synthesis of the building blocks themselves, since such design is outside the province of Boolean algebra.

Although it is not necessary to inclued a complete discussion of the Boolean algebra here, since the literature contains a number of such presentations,[1,2] a brief review of the basic principles may be helpful. The variables of the algebra represent leads or other entities, and can assume only two numerical values corresponding to the two states of these entities; at any given time a variable is equal to either 1 or 0: 1 if the corresponding lead is in an energized, or active, condition, and 0 if the lead is in an unenergized, or passive condition. There are three operations available in the algebra: addi-

---

Paper 53-216, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 17, 1953; made available for printing April 23, 1953.

S. H. Washburn is with the Bell Telephone Laboratories, Inc., New York, N. Y.

The author acknowledges the very considerable part played by W. Keister of the Bell Telephone Laboratories in the formulation of certain of the techniques presented in this paper. In addition, thanks are tendered to other members of the Bell Telephone Laboratories, especially J. H. Felker and W. D. Lewis, whose independent work in defining electronic building blocks and in suggesting means for interconnecting these blocks to satisfy circuit requirements has been of assistance in directing the thoughts of the author, and A. E. Anderson and R. M. Ryder, whose investigation of switching circuit design principles helped to motivate the work leading to this paper.

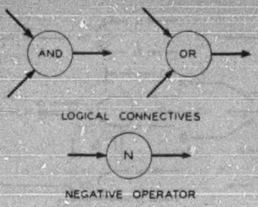


Fig. 1. Symbolic logical operators

tion (+), multiplication (·), and negation ( )′. The addition and multiplication serve to link the variables (or groups of variables already linked), and the negation, or prime, operates on a single variable or group of variables. In the convention to be adopted here multiplication is equivalent to the logical connective AND, and addition is equivalent to the logical connective OR. The operation of negation serves to transform the numerical value of the variables or expression on which it operates; that is, if $X=0$, $X'=1$ and, if $X=1$, $X'=0$; where X can be any algebraic function. It can be shown that the negative of any function can be derived from that function by replacing all products by sums and sums by products, and substituting for each variable its negative, noting that $(X')'=X$.

Fundamentals of Graphical Symbolism

The Boolean algebra is capable of indicating any logical relationship expressible by negation and the connectives AND and OR. Thus, if the requirements for, say, energizing a particular lead in terms of the energization and non-energization of other leads are expressed in these terms, an algebraic equivalent of this situation can be written.

Now, a graphical symbolism may be used which is equivalent to the logical statement, and to the corresponding algebraic expression. The components of this symbolism, shown in Fig. 1, are similar to the neuron symbolism used by J. Von Neumann and others.[3] Each symbol represents a logical operation performed on one input function, as in the case of negation, or on two input functions, as in the case of the connectives AND and OR. These symbols are interconnected in accordance with the logical structure of the particular algebraic expression to be represented. The final configuration is then a diagram graphically indicating the specific operations and their relationships required in the original logical statement.

A number of examples showing symbolic configurations and their corresponding algebraic expressions are shown in Fig. 2. In the first, the algebraic expression represents the statement "X is energized at all times that C is energized, or that A and B are energized"; the diagram shows that A and B are combined by the AND operator and the result is combined with C by an OR operator. In the second illustration, the verbal expression represented is: "X is energized at all times that C is energized together with either B energized or A not energized." Here, the diagram shows that B and the negative of A are combined by OR, and the result is combined with C by the operator AND. The remaining examples can be similarly interpreted.

Since algebraic expressions can be manipulated through the use of the theorems, many equivalent configurations can be drawn to represent equivalent logical statements. For example, consider the statement: "X is energized at those times when B and C are energized or when A or B, but not both, are energized. This is expressed algebraically as

$$f(X)=BC+(A+B)(AB)' \qquad (1)$$

since "A or B, but not both" is equivalent to "A or B and not A and B". This expression may be manipulated at will

$$\begin{aligned} f(X)&=BC+(A+B)(A'+B') \\ &=BC+AB'+A'B \\ &=AB'+B(A'+C) \end{aligned} \qquad (2)$$

The graphical symbolism for these is shown in Fig. 3.

It is interesting to note that either one of the two connective symbols is not strictly necessary in representing an algebraic expression. This follows from the fact that

$$AB=(A'+B')'$$

and that

$$A+B=(A'B')' \qquad (3)$$

It appears that any AND symbol can be replaced by an OR and three negation symbols, and that any OR symbol may be similarly replaced by an AND and three negation symbols. These and other equivalences derived by similar reasoning are shown in Fig. 4. The final equivalence in this figure, showing that two negation operators in tandem are equivalent to the absence of an operator, that is, a direct connection, is justified by the theorem

$$(X')'=X \qquad (4)$$

By applying these equivalences, various equivalent graphical configurations can be derived directly.

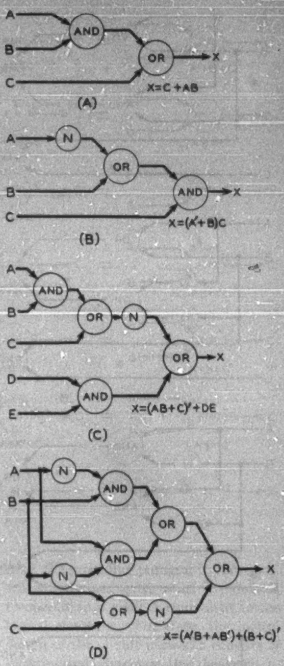


Fig. 2. Typical symbolic configurations

Application to Electronic Switching Circuits

The basic circuit blocks with which electronic switching circuits are constructed are switching gates, inverters, and memory elements. The gates may be considered as operators similar to the logical connectives, each gate effecting the energization of an output lead depending on the state of its input leads. The inverter is analogous to the logical negation operator. Memory elements, to be considered briefly later in this paper, are two-state devices having the property that, once driven into an active state, an element will remain in that state for some given interval.

In electronic switching circuitry, the difference between energization and non-energization is generally one of potential, although it may be one of current flow. That is, the lead may be at either one of two potentials, one more positive than th

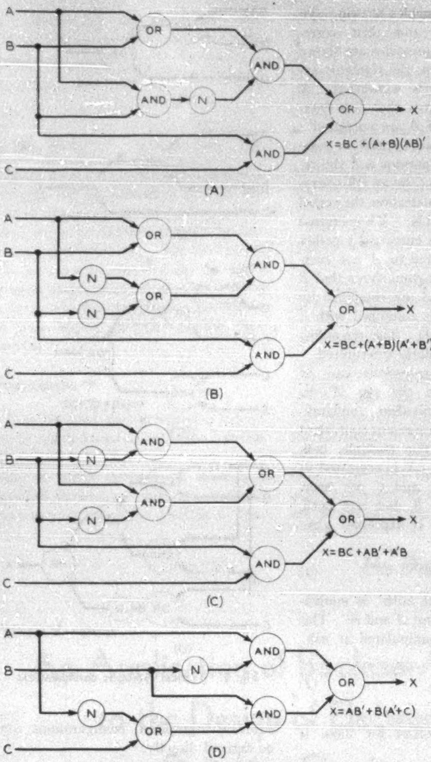


Fig. 3. Equivalent symbolic networks

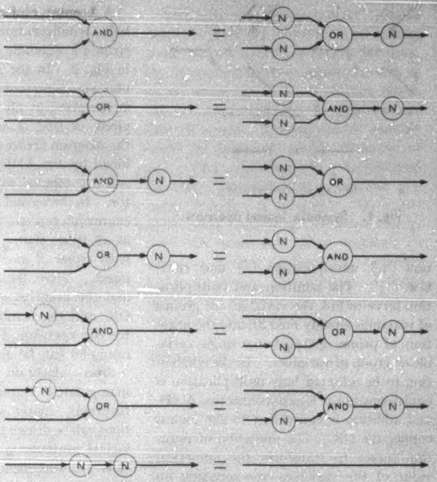


Fig. 4. Table of logical equivalents

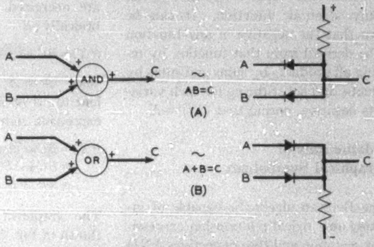


Fig. 5. Simple AND and OR gates

other. As a convention in this paper, these two potentials will be arbitrarily designated "+" and "−," the "+" potential being the more positive of the two. The amplitudes of the two potentials, and also the fact that the potential designated "−" may actually be positive with respect to ground, is of no concern here.

These indications of energization and nonenergization may be used to advantage with the symbolism just described, to clarify the analysis of a symbolic configuration and to specify the operation of the individual gate elements. For example, a particular AND operator, or gate, may have the characteristic that when, and only when, both input leads are "+," the output lead is "+". This gate can be represented as in Fig. 5, with a "+" symbol associated with each input lead and with the output lead. Similarly, the OR gate in the same figure has "+" signs associated with all its leads, indicating that the output lead is "+" when either or both input leads are "+." Typical circuits which operate according to these specifications are also shown in the figure.

Now, for a particular case, either the "+" potential or the "−" potential may be the energized state. For convenience in adopting the Boolean algebra to electronic circuitry, the "+" and "−" potentials will be considered as represented by the numerical constants 1 and 0, respectively. Therefore, the algebraic expressions which describe the operation of the AND and OR gates of Fig. 5 are, in that order

$$AB=C \text{ and } A+B=C \qquad (5)$$

The operation of these gates can be described by an alternative form. For example, for the first gate of Fig. 5, the operation of the gate can be described by the phrase "the output is "+" when, and only when, both inputs are "+"," or by the phrase "the output is "−" when either input is "−"." This, of course, follows from the equivalence of the two algebraic expressions

$$AB=C \text{ and } A'+B'=C' \qquad (6)$$

Both expressions represent exactly the same logical situation. Similarly, since $A+B=C$ and $A'\,B'=C'$ are equivalent to one another, the behavior of the second gate may be stated: "the ouput is "−" when, and only when, the inputs are both "−"." As a result of these equivalences, two different symbols can be used to represent the same operational characteristic, as indicated in Fig. 7.

The "+" and "−" labels may also serve to stress the behavior of the negation operator, or inverter. When the input lead to the inverter is "+," the output lead is "−" or, equivalently,

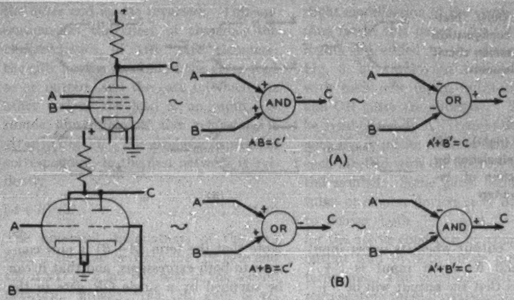


Fig. 6 (above). Gates performing inversion

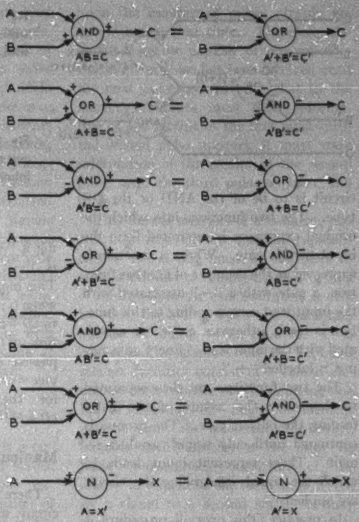


Fig. 7 (right). Table of gate elements for switching circuit application

when the input is "−," the output is "+." As a result, the two symbolic forms shown in Fig. 7 obtain.

Gates and inverters can be combined exactly as were the logical connectives and the negation operator in Fig. 4. However, it is often more convenient to consider the combinations thus obtained as new elemental building blocks rather than as interrelations of the original elements, since in many cases the basic electronic devices used in physical circuits can be described directly by these new blocks. As an illustration, consider the pentode of Fig. 6. With suitably chosen grid biases, a "−" signal on either grid will drive the tube to cut-off; both grids must be "+" to allow the plate potential to fall. The symbolic equivalents for the element and their algebraic expressions are shown in the figure. A second illustration is the pair of plate-connected triodes also shown in Fig. 6. Here, "+" potentials must be applied to either grid to cause the plate potential to assume the "−" potential. Therefore, the element is represented by either of the symbols shown in the figures.

Considering elements of this type as variations of switching gates, it is clear that there are, at most, six operationally unique gates, corresponding to the six logically different combinations of Fig. 4. Four of these have already been introduced. Each gate, as has been seen, can be represented by a pair of equivalent symbols and by the corresponding pair of characteristic equations which define the symbols. These symbolic representations together with their algebraic expressions, as well as those for inversion, are given in Fig. 7.

Synthesis of Switching Networks

The networks to be discussed first are of the type known as combinational circuits. These are circuits in which the state of the circuit output at any time is completely determined by the state of the circuit input at that time. The particular time sequence in which various signal conditions are applied to the circuit input has no effect on this dependency. The design of circuits in which operation is affected by the sequence of application of input signals is considered later in this paper.

A circuit network to produce a given output signal corresponds to a logical statement written in terms of the AND and OR relations between control conditions. It has been seen that such a statement may be represented graphically by a configuration of AND, OR, and negation elements suitably interconnected. Also, the preceding paragraphs have presented switching gates which are defined to operate in a manner equivalent to the logical elements. Therefore, the logical design of the circuit networks consists of selecting and interconnecting these switching gates to correspond to the statement describing the relationship between the output signal and the control conditions.

The technique for carrying out the design procedure is a routine one. First, the circuit requirements are written as an algebraic expression. Any complex expression in Boolean algebra may be separated into two functions related by either multiplication or by addition. This determines whether the gate whose output lead is the output lead for the

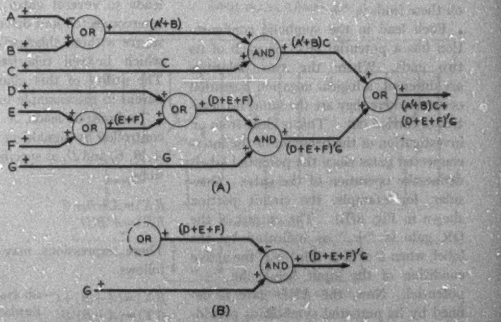


Fig. 8. Typical switching network configuration

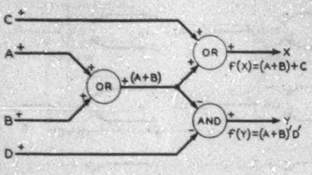


Fig. 9 (left). Network configuration with common circuit elements

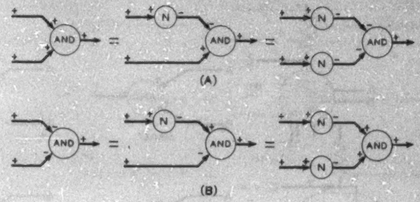


Fig. 10 (right). Circuit manipulation by introduction of inverters

circuit is to be of the AND or the OR type. The two functions into which the original expression is separated form the input to this gate. Where a function is expressed as the negative of another function, a gate with a "−" associated with the input lead corresponding to this function is used; otherwise a "+" is associated with the input lead. Every gate output is labelled "+."

The two functions are then separated to determine the nature of the gates feeding the output gate. The process is continued until only single variables remain. These represent input leads to which the signals controlling the circuit are applied.

In practice, this over-all procedure is very simple—merely a translation from algebraic expression to configuration of gates, replacing each multiplication by an AND gate, each addition by an OR gate, and translating each negation symbol to a "−" sign on the requisite gate input lead. The process is illustrated in Fig. 8(A) for the expression

$$(A'+B)C+(D+E+F)'G \qquad (7)$$

In the figure, the algebraic expression for the signal on each lead is indicated. Input leads for the functions A' and $(D+E+F)'$ have been labelled "−" to provide for the required negation. The assumption that each gate element has only two inputs makes it necessary to produce the expression $(D+E+F)$ in two stages, that is, $D+(E+F)$. Note that the potential labels on the circuit input leads indicate that the active signal on these leads is "+."

Each lead in the symbolic configuration has a potential label on each of its two ends. Where the two polarities are different, a logical negation operation occurs; where they are the same, no negation is carried out. This is clear from an investigation of the behavior of the interconnected gates since the potential labels define the operation of the gates. Consider, for example, the circuit portion shown in Fig. 8(B). The output of the OR gate is "+," as indicated by the label, when D or E or F is "+," the active condition of the input being the "+" potential. Now, the AND gate is defined by its potential symbols as providing a "+" output when its upper input is "−" and its lower input is "+." This means that the output will be "+" when G is "+" and the lead corresponding to $(D+E+F)$ is "−." The active condition at the output of the OR gate is the passive condition for the AND gate and vice versa. Thus, the active condition for the AND gate corresponds to $(D+E+F)'$.

Manipulation of Gate Networks

There may be a number of different circuit arrangements corresponding to a given logical expression. It is often desirable to investigate several of these to determine the one most suitable for a given application. Rearrangement of a circuit without changing its over-all operating characteristics may be accomplished either by manipulation of the corresponding algebraic expressions, in accordance with the fundamental theorems, or by certain transformations which may be made directly on the circuit. Unfortunately, the Boolean algebra in its present state of development can specify no explicit set of operations which will lead to the most economical circuit form.

A fortunate characteristic of the circuitry composing gate elements is the virtual isolation between input leads. That is, a signal or change of signal on one input lead has little or no effect on the other input lead. As a result, input leads to several gates may be directly interconnected and driven from the same source without the fear of back-up paths which bedevil relay circuit designers. The utility of this characteristic is apparent in the example to follow.

Two output leads, X and Y are to be controlled by signals on four input leads, A, B, C, and D, as specified by the expressions

$$f(X)=A+B+C$$
$$f(Y)=A'B'D' \qquad (8)$$

These expressions may be modified as follows

$$f(X)=(A+B)+C$$
$$f(Y)=(A+B)'D' \qquad (9)$$

Note that the term $(A+B)$ is now common to both expressions, and that it can be supplied by a single OR gate in the circuit configuration. The complete network is shown in Fig. 9.

In a practical case, the physical manifestations of certain of the six gates symbolized in Fig. 7 may not be economical or reliable, and the use of these gates may therefore be prohibited. Such gates can be avoided by algebraic manipulation or by the introduction of inverters. The effect of an inverter in the input lead to a gate is that of interchanging active and passive states on that lead or, in terms of the symbolism, of interchanging "+" and "−" labels. This is shown in the typical equivalents of Fig. 10. Note that comparison of the potential symbols at the two ends of each lead still indicates whether or not a logical negation is taking place.

Another technique for avoiding undesirable types of gate elements is that of reversing potential symbols on interconnecting leads. In any network, the potential symbol at the one end of a given lead may be reversed if the symbol at

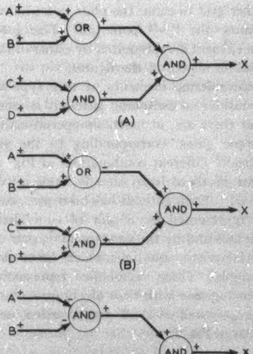


Fig. 11. Circuit manipulation by reversal of potential symbols

the other end is also reversed. This operation has the effect of changing the characteristics of two gate elements: the gate having one end of the given lead as an output, and the gate to which the other end of the lead is an input. A simple illustration of this relabeling is given in Fig. 11. The algebraic expression corresponding to the first network is evidently

$$(A+B)'CD \qquad (10)$$

The potential designations of the two ends of the lead between the OR gate and the AND gate are reversed in the second network. Finally, the OR gate is replaced by the equivalent AND gate in the third network. The network is operationally equivalent to the original network since

$$(A+B)'CD=(A'B')CD \qquad (11)$$

Omission of Gates

In some instances, a gate through which there is no inversion can be replaced by a junction point. These gates are of the characteristic types $A+B=C$, $A'+B'=C'$, $A'B'=C'$ and $AB=C$. Whether such a gate can be omitted in a particular case depends upon the signal levels and biases associated with the input and output leads, and also upon whether direct connection of these leads at a junction point will introduce unwanted signal conducting paths.

Other Types of Gates

Gates having more than two input terminals and one output terminal require slight modification of the symbolism used thus far in this paper. If, for example, inputs are added to the diode gates of Fig. 5, by employing a diode in series with each new input, multi-input AND and OR gates are obtained. These may be symbolized by AND and OR elements with the requisite number of input leads, each labelled with the appropriate potential symbol. Some gates using vacuum tubes or similar devices may be arranged to deliver both "+" and "−" signals for the active condition by connecting an output lead to both anode and cathode terminals. The symbolic gate elements representing these have two output leads, one labelled "+" and the other labelled "−." The techniques for dealing with these elements are exactly the same as for those elements already discussed.

Synthesis of Sequential Circuits

A sequential switching circuit is a circuit which assumes a number of states, each dependent not only upon the particular input signals, or combination of input signals, applied but also upon the sequence in which these signals appear. In other words, the behavior of a sequential circuit, upon receipt of a given input signal, depends upon its state resulting from previous input signals. This is in contrast to the combinational-type circuits which have just been discussed. The essential difference between these two types of circuit lies in the requirement that the sequential circuit, unlike the combinational circuit, must include some form of memory to make possible a differentiation between similar signals separated in time.

A circuit designed to count a number of identical pulses is a familiar example of a sequential circuit. As pulses are received, the counting circuit passes from one state to another. The state of the circuit at any instant during the process depends upon the number of pulses already received before that time. Therefore, the circuit must be capable of remembering that a given pulse has been received until the next pulse has arrived; that is, the memory facilities must store information of the receipt of the first pulse until the second pulse appears, it must store information of the receipt of the second pulse until the third pulse appears, and so on.

This partial analysis of a counting circuit indicates that the memory capabilities of the circuit must be sufficient to store a number of different pieces or bits of information, the exact number, as will be seen, depending upon the maximum number of pulses to be counted. It is convenient to assume that the required storage facilities consist of 2-valued elements, each capable of remembering one bit of information. Such an element may be a gas tube, semiconductor device, bistable multivibrator, or the like; its basic characteristic consisting of the ability to remain in one state until driven by a signal to a second state. It may then remain in this second, active state until cleared by a second signal or until, after a predetermined time has elapsed, it returns to its original state under its own control. Individual cells on magnetic drums and memory tubes have simi-

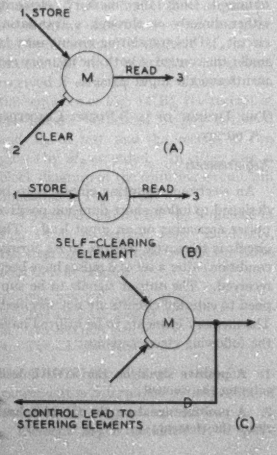


Fig. 12 (left). Typical memory elements

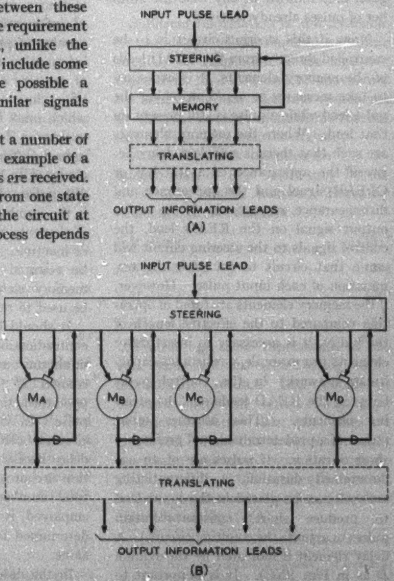


Fig. 13 (right). Block diagrams of counting circuit lar characteristics, and may also be considered as separate memory elements.

To generalize, a memory element may be indicated in symbolic form as shown in Fig. 12(A). A signal on lead 1 (STORE) causes the element to assume its active state. A signal on lead 2 (CLEAR) clears the element, restoring it to its original passive state. An indication of the state of the element is obtained from lead 3 (READ). It is convenient to assume that this latter information is continuously available, although this is not necessary. Where the element is self-clearing, maintaining an active state for only a limited time, the CLEAR lead is omitted, as in Fig. 12(B).

Returning now to the analysis of the counting circuit requirements, it is clear that if the input pulses are themselves to actuate the various memory elements, some form of steering is necessary between the input pulse lead and the memory elements. If, for example, the first pulse is to actuate a memory element M_A and the second pulse is to actuate a memory element M_B, the first pulse should be directed to the STORE lead of element M_A and the second pulse to the STORE lead of element M_B. This steering function associating the input pulse lead with the memory elements can be under the direct control of these elements since, taken in aggregate, they indicate at any time the number of pulses already received.

Now, if this steering circuit is to be controlled directly from the READ leads of the memory elements, it is necessary to take measures to avoid switching the pulse lead while a pulse is still present on that lead. Where the memory elements are such that there is sufficient delay between the appearance of a STORE or CLEAR signal and the appearance and disappearance, respectively, of the active output signal on the READ lead, the control signals to the steering circuit will cause that circuit to act after the termination of each input pulse. However, if the memory elements are rapid in operation compared to the effective length of the pulses, it is necessary to insert delay elements (for example, a resistance-capacitance network) in the control paths between the READ leads and the steering elements. (This assumes input pulses of a predetermined, and preferably short duration. If pulses are of an undetermined duration, a differentiating network can be placed in the input lead to produce short, constant-duration pulses to actuate the counting circuit.) A delay element is indicated by the symbol D as in Fig. 12(C). It is important to note that this delay is not included in the memory requirements of the over-all circuit, but is necessitated by the steering function between the input pulse lead and the memory elements.

The requisite output indications to be passed to external circuits may be obtained from the memory elements. The output indications may be derived directly from the READ leads of the memory elements, or, if necessary, the READ leads may actuate a translating circuit to provide the count information in the desired form.

The over-all counting circuit may be shown in block form as in Fig. 13(A), to illustrate the required functions to be carried out in the circuit. Separating the memory block into its component parts, the symbolic representation of the circuit is as indicated in Fig. 13(B).

MEMORY REQUIREMENTS

In the design of a circuit capable of counting a certain maximum number of pulses, it is convenient to specify the minimum number of memory elements which must be provided. This may easily be determined from a consideration of the quantity of information which must be stored. Briefly, the circuit at any time must know, or must have remembered, how many pulses have been received. Thus, the circuit must have a capacity for storing a number of independent pieces of information equal to the maximum number of pulses. Thus, if these bits of information are stored in the collection of memory elements on a combinational basis, the number of memory elements, N, which must be provided is given by the expression $P \leq 2^N - 1$, where P is the maximum number of pulses, or by $P \leq 2^N$ if the circuit is to be restored to normal by the Pth pulse. It will be recognized that 2^N is the number of unique combinations of N elements, each of which is either active or inactive. For example, 15 pulses can be counted by a circuit containing 4 memory elements; the 16th pulse may be used to restore the circuit to normal.

It should be noted that this is not in contradication to the expression for the minimum number of relays R, in a relay counter: $2P \leq 2^R - 1$ or $2P \leq 2^R$, depending upon the action of the circuit after the Pth pulse. In the case of relay sequential circuits, both the memory facility and the delay necessitated by the steering function are provided by relays. Thus, the total number of relays which must be employed is greater than the number determined by the memory requirement alone.

In the design of a counting circuit, of course, the number of memory elements may exceed the theoretical minimum. This comes about because, in some cases, the arrangement used to steer the pulse lead may be simplified by the use of additional memory elements.

GENERAL DESIGN ROUTINE

After determining the number of memory elements to be utilized, an operating sequence is specified for these elements. This process corresponds to the formulation of a sequence diagram for a sequential relay circuit. By inspection of this sequence diagram, the requirements for actuating and clearing each memory element may be stated in terms of the input pulses and the states (active or passive) of all the memory elements. These requirements indicate the configuration of a suitable steering arrangement composed of gate elements; this circuit may be designed by the routine method already presented. Finally, if necessary, a translating circuit is similarly designed to provide the requisite output indications.

Although the preceding paragraphs have presented the basis for the design of a circuit for counting pulses, it is evident that the method is applicable to the design of any sequential circuit composed of gate and memory elements. In the general case, the single pulse lead of the diagram of Figure 13(B) is replaced by a number of input signal leads. Sufficient storage capacity is provided in accordance with the output requirements of the circuit, and a steering circuit composed of gate elements is placed between the input signal leads and the memory elements. The output signals may be obtained from the memory elements either directly or through a translating circuit. This translating circuit may be under the control of both the memory elements and the input signals.

THE DESIGN OF A 3-PULSE COUNTING CIRCUIT

Requirements

An electronic counting circuit is to be designed to follow short-duration positive pulses appearing on an input lead. The circuit is to recycle (restore to its normal condition) after a set of 3 pulses have been received. The output signals to be supplied to external circuits are not specified. The memory elements to be utilized have the following characteristics:

1. A positive signal on the STORE lead actuates the element;
2. A positive signal on the CLEAR lead clears the element;

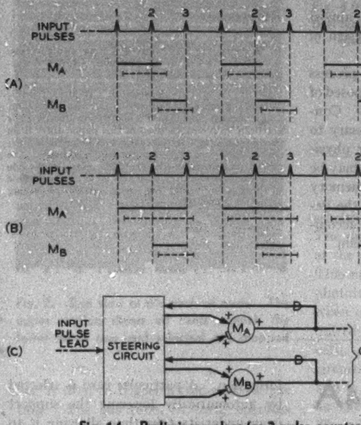


Fig. 14. Preliminary plans for 3-pulse counter

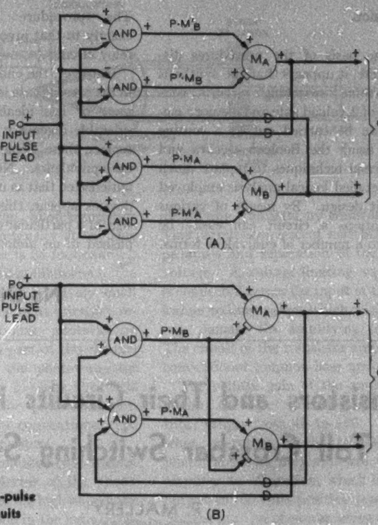


Fig. 15 (right). Three-pulse recycling counting circuits

3. A positive signal on the READ lead indicates that the element is in an active condition.

Solution

Since the number of pulses to be counted, 3, is greater than 2^1 but less than 2^2, it is evident that at least 2 memory elements must be provided. It appears that several sequences for the operation of these elements, M_A and M_B, are possible; two of these are indicated in Figs. 14(A) and 14(B). In both diagrams, the solid lines indicate the periods of activation of the memory elements, and the dotted lines indicate the resultant signals passed to the steering circuit through the delay elements. In the first of these sequences, the counting circuit knows that one pulse has been received if element M_A is actuated, and that two pulses have been received if element M_B is actuated. In the sequence of Fig. 14(B), the receipt of one pulse is indicated by M_A actuated and M_B at rest, and the receipt of two pulses by M_B actuated. In either case, a block diagram of the counting circuit will appear as in Fig. 14(C).

A symbolic circuit for the sequence of Fig. 14(A) is derived first. Evidently, the steering circuit must switch the pulse lead to four points:

1. The STORE lead of M_A;
2. The CLEAR lead of M_A;
3. The STORE lead of M_B;
4. The CLEAR lead of M_B.

Using switching algebra, a steering arrangement is determined by inspection of the sequence diagram, as follows:

Element M_A is actuated by a pulse occurring when both M_A and M_B are unactivated. Thus, the algebraic expression indicating the energization of the STORE lead of M_A is

$$f(M_A)_S = PM_A'M_B' \qquad (12)$$

Element M_A is cleared when M_B is in an activated state and no pulse is present. Therefore

$$f(M_A)_C = P'M_B \qquad (13)$$

Element M_B is actuated by a pulse occurring when M_A is actuated (as indicated on the lead from the READ lead of M_A through the delay D). Therefore, the requirement for energizing the STORE lead of M_B is

$$f(M_B)_S = PM_A \qquad (14)$$

Element M_B is cleared by a pulse occurring when M_A is at rest and M_B is actuated. Thus

$$f(M_B)_C = PM_BM_A' \qquad (15)$$

Further inspection of the diagram of Fig. 14(A) indicates that the above expressions for $f(M_A)_S$ and $f(M_B)_C$ may be simplified. Since the STORE lead of M_A may be energized while M_A is in an actuated state, without affecting the state of M_A, both the first and second input pulses may be placed on the STORE lead of M_A. Therefore

$$f(M_A)_S = PM_B' \qquad (16)$$

Similarly, a signal may be placed on the CLEAR lead of M_B when M_B is unactuated without altering its state; both the first and third pulse may appear on the CLEAR lead of M_B. Therefore

$$f(M_B)_C = PM_A' \qquad (17)$$

Sufficient information is now available for the design of the required switching circuit. By inspection of the algebraic expressions 13, 14, 16 and 17 above, and by translating the algebraic symbolism to the gate-element symbolism, the circuit of Fig. 15(A) may be drawn.

A circuit which operates on the sequence plan of Fig. 14(B) may be derived in a similar manner. Here, by inspection of the diagram

$$f(M_A)_S = P \qquad (18)^*$$

$$f(M_A)_C = PM_B \qquad (19)$$

$$f(M_B)_S = PM_A \qquad (20)$$

$$f(M_B)_C = PM_B \qquad (21)$$

Translating these expressions to the gate symbolism, the circuit of Fig. 15(B) immediately results. This circuit is obviously preferable to that of Fig. 15(A) in that two, rather than four, gates are employed. That this would result might have been anticipated since the sequence diagram of Fig. 14(B) evidently requires a less complicated steering procedure than that of Fig. 14(A).

---

* Note that, in this expression, use is made of the assumption that a CLEAR signal overrides a STORE signal to a memory element. If this cannot be assumed, $f(M_A)_S$ could be PM_A'.

Conclusion

On the basis of the procedures discussed here, it appears that the synthesis of electronic switching circuits constructed of 2-valued gate and memory elements can be carried out in a routine manner using the Boolean algebra and other formal techniques to replace much of the detailed logical analysis employed in circuit design. By means of various manipulations a circuit can easily be altered to a number of equivalent forms. The procedure corresponds almost exactly to that involved in the design of relay circuits.

However, the end result of this process in the present case is a circuit composed of more or less idealized elements. Considerable ingenuity may be necessary to replace these elements with their physical equivalents. Nevertheless, it may be anticipated that as new gates and memory devices appear, this final step in the design of a particular circuit may be accomplished in an almost routine fashion.

References

1. A Symbolic Analysis of Relay and Switching Circuits, **Claude E. Shannon.** *AIEE Transactions*, vol. 57, 1938, pp. 713–23.

2. The Design of Switching Circuits (book), **Keister, Ritchie, Washburn.** D. Van Nostrand and Company, Inc., New York, N. Y., 1951. (The algebraic conventions used in this paper differ from those found in this reference and from those normally employed by the author. However, the present conventions are probably more applicable to electronic circuitry, since it is more reasonable to think of high- and low-voltage signals as having values of 1 and 0, respectively.)

3. Calculating Instruments and Machines. **D. R. Hartree.** The University of Illinois Press. Urbana, Ill., 1949.

No Discussion

Transistors and Their Circuits in the 4A Toll Crossbar Switching System

P. MALLERY
NONMEMBER AIEE

NATION-WIDE dialing is one of the more recent advances in the telephone art and has been the subject of papers presented before the AIEE.[1] The heart of nation-wide dialing is the new no. *4A* toll crossbar switching system. It is this system which, at the various control switching points, receives the digits dialed by the toll operators or by the subscribers. It then automatically determines the routing of the call to any destination in the United States, as well as in Canada, and proceeds to establish the connection to succeeding control switching points or to the local office. In this system it is necessary to translate the digits received into information as to the location of trunks on the switches, the kind of outpulsing required, the number of digits to send forward, and other such instructions.

The translator needed for the no. *4A* system must have a large capacity, for it has to be capable of furnishing the information to route a call to any one of thousands of central offices in the United States and Canada. As frequent changes are made in the routing of calls due to the growth of the system and modifications of central offices, corresponding changes must be made in the routing information at the control-switching points. It is therefore necessary to be able to quickly and economically change the routing information pertaining to any given set of input digits. In short, the translator must perform functions similar in nature to those accomplished by quick reference card files.

In order to meet these requirements it was necessary to develop a new type of translator[2] radically different from any previously used. Its operation is based on selecting a particular card from a stack of cards in accordance with the input digits. It obtains the routing information from the output code registered on the card. To determine this output code the translator depends upon that revolutionary new device, the transistor. The purpose of this paper is to describe the transistor and its circuits, as used in the card translator.

Fig. 1 shows a card with the output information registered on it in the form of enlarged holes. In the translator this card stands in a vertical position, resting on the tabs along its lower edge. When all cards are in their normal position, they line up and the holes in the cards form unobstructed, horizontal tunnels, called channels, through the entire stack. The tabs on the lower edge of the card correspond to the input digits for which this card supplies the routing information. A particular card is selected by automatically lowering the support from under its tabs thus allowing it to drop a distance slightly greater than the height of an unenlarged hole. All channels through the stack of cards will then be blocked by the dropped card except those for which the holes in the card have been enlarged. This results in a pattern of clear channels which represents the output information. Determination of whether a channel is clear or blocked is made by passing light beams through the holes. Fig. 2 illustrates the change from all cards normal to one card dropped.

The Channel Circuit

Fig. 3 is a block diagram of the circuit used to determine whether a particular channel is interrupted by a dropped card or not. Each block, with the exception of the light source, represents a piece of equipment provided individually for each channel. The light source is common to all channels. If the hole in the dropped card for a particular channel has been enlarged, the light will pass completely through the stack of cards and fall on the phototransistor. The phototransistor converts the presence of light into an electrical signal which, after being increased by the transistor ampli-

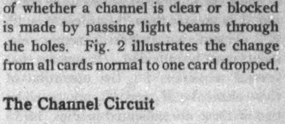


Fig. 1. Translator card showing tabs used for selection, and the enlarged holes representing routing information

---

Paper 53-218, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 17, 1953; made available for printing April 23, 1953.

P. Mallery is with the Bell Telephone Laboratories, Inc., New York, N. Y.

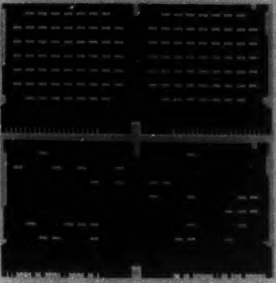


Fig. 2. End view of the stack of cards. The upper section shows all cards normal, the lower shows one card selected and dropped

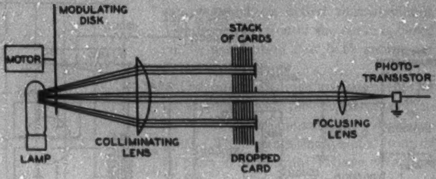


Fig. 4 (right). Optical section of channel circuit

fier, is used to trigger a cold-cathode gas tube. The gas tube, in turn, operates the channel relay. This relay is located in the associated equipment which uses the information supplied by the translator to actually process the call.

In making a detailed examination of the channel circuit it is convenient to break it down into two parts, the optical section and the electrical. The optical section includes everything up to the point where the light falls upon the germanium of the phototransistor. This part of the channel is shown functionally as Fig. 4. The light source is a standard projection-type lamp normally rated at 500 watts. To obtain long life it is operated in the translator at half voltage at which level its input is approximately 170 watts. This type of lamp was chosen because of its high concentration of light in a small plane.

The light from the lamp passes through a motor-driven perforated disk which modulates it with an approximate square wave at a 400-cycle rate. Modulated light is used because it is easier to build a-c than d-c amplifiers, and also, in this manner, the ratio between the light and dark currents of the phototransistor is the important factor rather than the absolute value of either. The modulated light is collimated by a lens to minimize the loss as the beam passes through the holes in the card. Unless interrupted by a dropped card, this light beam will pass through all of the cards and fall on the lens which focuses the light on the sensitive area of the phototransistor. The minimum light intensity at the lens of the phototransistor is 34 foot-candles. This supplies about 12 millilumens to the phototransistor, a figure relatively small when compared to the light intensity required by conventional photoelectric cells. The electrical part of the channel circuits starts with the phototransistor and is shown in Fig. 5. The light acts as the emitter of the phototransistor. The collector is of the conventional type for point-contact transistors.

As is normal in grounded-base transistor circuits, the collector of the phototransistor is biased in the high impedance direction. A variation in the light intensity causes a variation in the collector impedance of the phototransistor. The type used has an impedance of about 10,000 ohms when dark, which is reduced to approximately 3,000 ohms when illuminated. The output of an illuminated phototransistor, when coupled to the amplifier, ranges from 1.3 to 12 volts, positive peak (400 cycles), depending upon the age and condition of the transistor.

Since the discrimination by the channel circuit between a clear or blocked light path depends upon the presence or absence of an a-c output from the phototransistor, noise of sufficient magnitude, if present when the channel is dark, would cause a false indication. To guard against such false operations the Western Electric Company, during the manufacturing process, checks each phototransistor for dark noise. If during a 5-minute interval the dark voltage exceeds 20 millivolts, that phototransistor is rejected.

The phototransistor is coupled to the amplifying transistor by transformer *T1*. This permits convenient matching of impedances and separation of the d-c bias voltage. A voltage-limiting varistor *V* is connected across the input of the transformer to limit surges which might otherwise damage the amplifying transistor. The circuit of the transistor amplifier is a conventional common-base arrangement.

The voltage gain of the amplifier, including the input transformer to the gas tube, ranges from 40 to 100. However when operating in the translator, the phototransistors normally will drive the amplifier to saturation, which limits the output to 160 volts positive peak or less. For the purpose of guaranteeing operation a minimum output voltage of 38.5 has been set as a rejection point for a phototransistor-amplifier combination.

The output of the transistor amplifier is normally sufficient to break down the control gap of the *376B* cold-cathode gas tube. Sufficient current flows in this control gap to insure reliable transfer to the main gap when the output control relay *O* in the associated equipment operates. But to provide greater operating margin, the bias of −24 volts is removed from the control anode just before channel operation is required. The relay *R*, which substitutes ground for the bias voltage is operated by a circuit which checks that the card being dropped is completely down. This down-check circuit utilizes two holes in the card which are never enlarged and employs two photo transistors to detect the presence or absence of light through these holes.

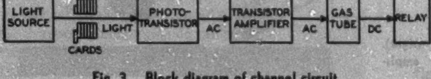


Fig. 3. Block diagram of channel circuit

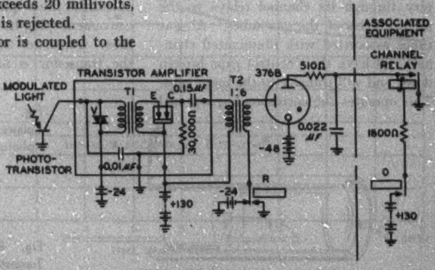


Fig. 5. Electrical section of channel circuit

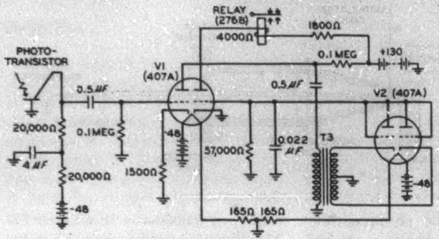


Fig. 6. Card-down check circuit

Fig. 6 shows the card-down check circuit. It differs from the routing-information channel circuits in that the operation of a relay is required when the light is blocked. Therefore the transistor amplifier-gas tube circuit is not applicable. For the down-check channels, the output of the phototransistors is amplified by the first section of a conventional double-triode thermionic-emission tube *V1*. The a-c signal at the plate of *V1* is rectified by a conventional full-wave rectifier consisting of transformer *T3* and tube *V2*. The rectifier voltage is negative with respect to ground and when it is impressed on the grid of the second section of tube *V1*, that section is driven past cutoff. Therefore, as long as light falls on the phototransistor, no current flows through the relay. When the dropped card blocks the light, the negative voltage on the grid disappears and the tube conducts, thus operating the associated relay. While the prime purpose of this down-check is to signal the associated equipment that the card is in position for recording its output code, advantage is also taken of this circuit's ability to distinguish between a light and dark phototransistor at any time to provide alarms in the event of a lamp or modulating disk failure.

After the card has been checked down and the associated equipment is ready to accept the output of the card, that equipment connects a positive 130-volt battery through its channel relays to the main anodes of the gas tubes. All gas tubes associated with illuminated channels will have their control gaps broken down and will transfer to the main gaps. This operates the corresponding channel relays in the associated equipment. The relays, in operating, lock to ground and thereby extinguish the main gap discharges, thus increasing the life of the gas tubes. Those channels which have been blanked out will not have the control gaps of the gas tubes broken down. Therefore when the 130 volts is applied, the relays associated with the darkened channel will not operate. The operation or nonoperation of the relays in the associated equipment completes the function of the channel circuits.

The capacitor and resistor network at the main anode of the tube is to prevent transients due to the operation of other channels from falsely breaking down the main gap of a dark channel.

Channel Packages

The phototransistor is mounted in a tube along with a lens which focuses the colliminated light on the transistor. Fig. 7 is a cutaway view of the *3A* phototransistor showing the relationship of the lens to the transistor. To mount in the translator, the tube is slipped into an accurately positioned hole and clamped in place, using the slotted ear. This mechanical fastening is also the ground connection. The output lead from the collector is attached with the use of a slip-on connector.

The amplifying transistor, transformer *T1*, varistor *V*, the resistor, and two capacitors of the amplifier are packaged as a convenient plug-in unit. Fig. 8 shows these transistor amplifiers. As shown the transistor is mounted under a removable cap on the package so that it may be conveniently replaced if necessary.

The gas tube, transformer *T2*, and the associated resistor and capacitor are also assembled as a packaged unit.

Transistor Performance Tests and Data

Like all classes of devices, transistors have certain characteristics which determine their suitability for any given application. In the card translator, the size advantage as compared to electron tubes was, of course, valuable. However, their ability to work at very low power levels and their relatively long life were the most important factors in their selection. Exactly how sensitive they were and what was their life had to be determined. Also transistors are not free from noise and, in addition, they are known to be affected by temperature. The effects of both these parameters on the ultimate performance of the card translator also had to be evaluated.

The reliability of the transistor circuit was the first and most important question to be answered. To test for this, three card translators were installed in the laboratory and equipped with a total of 272 channel circuits. These translators were connected in the normal service manner to standard associated *4A* crossbar system equipment. Then call after call was sent through the system. In all, 28,000,000 translations were made, each translation requiring the operation of many channels. Only one failure to operate occurred during these translations which might be ascribed to a failure of a channel circuit. The reliability of the circuit was thereby well established.

During the time the tests for reliable operation were continuing, the effect of the internal noise of the two types of transistors was being investigated. Noise produced by the phototransistor would be the more critical of the two as it would be amplified. If the noise was great enough, false operation of a channel relay would result. No false operations due to transistor noise were observed. To obtain data on whether there was margin against the

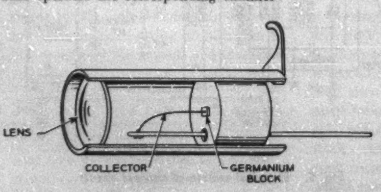


Fig. 7 (left). Cutaway view of phototransistor

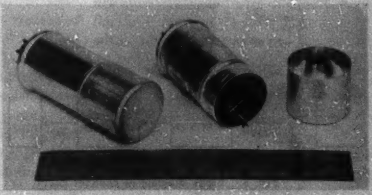


Fig. 8 (right). Transistor amplifiers

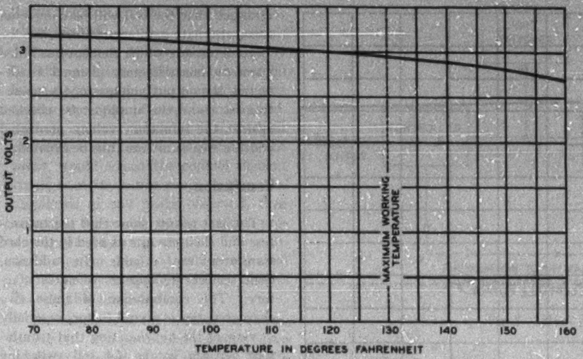


Fig. 9. Average temperature-output voltage characteristic of a phototransistor

effect of noise, several phototransistors which normally would be rejected by the Western Electric Company because of their noise voltage were installed in the translator. There was no adverse effect.

Temperature is one of the important parameters which affects the performance of transistors. Since the translator has powerful electromagnets working in close proximity to the phototransistors, there is a distinct heat rise, which had to be carefully considered. The *3A* phototransistor has an upper limit for reliable operation of 130 degrees Fahrenheit. This is not necessarily typical of other phototransistors and is lower than the temperature limit applying to the amplifying transistors. Fig. 9 graphs the change in output voltage of the *3A* phototransistor with respect to temperature. The output voltage of the transistor amplifier has approximately the same percentage change over the same temperature range. Tests for reliable operation were made on 96 of the channel circuits under temperature conditions corresponding to the maximum expected in service, and were found completely satisfactory.

While the reliable operation of the channel circuits under all conditions which might be encountered in service was of prime importance, the life of the transistors will have great effect on the cost of the apparatus and maintenance. Therefore 96 channels were subjected to a series of measurements designed to determine the effect of aging on the transistor elements. For this purpose a test set was used which was arranged to place the worst circuit condition on the transistor equipment that will occur in service. It was equipped to make output voltage measurements of the phototransistors and the transistor amplifiers.

The results of the series of output voltage measurements were subjected to the usual statistical analyzation method. This showed that the performance with age of the transistor elements fits the normal distribution curve. Fig. 10 is an average curve of the output voltage of a group of 96 phototransistors as they change with time. The months in operation represent the total integrated elapsed time during which the phototransistors were exposed to light with all voltages applied. The first group of points on the curve and the group ranging from the 6-month point on were obtained from two separate groups of phototransistors. The two solid lines are curves based on actual measurements. The dashed portion of the curve is based on the assumption that both groups age in the same manner.

The output voltage measurements of the amplifiers were the outputs of particular phototransistor-amplifier combinations. The average change of a group of 96 pairs in the output voltage with time is shown in Fig. 11. Again the first and last groups of points represent two different lots of transistor equipment.

Maintenance in Service

Although every effort has been made to insure the reliability of operation of the card translator and its associated equipment, it is not considered enough in the Bell System to depend solely on the reliability of equipment. It is important to determine, in advance of service failure, whether any element is becoming weak and to take corrective action before actual failure occurs. This is particularly true when devices which approach their end of life gradually are used. The transistor is such a device.

In the card translator this philosophy of preventing service troubles is carried out by periodically making a marginal test of the channel circuits. To accomplish this test the intensity of the light falling on the phototransistors is reduced and the associated equipment is used to check for channel operation. No card is dropped for this test; therefore no channel is blocked off and every channel relay should operate. A record is made by means of the associated equipment showing the channels which operated.

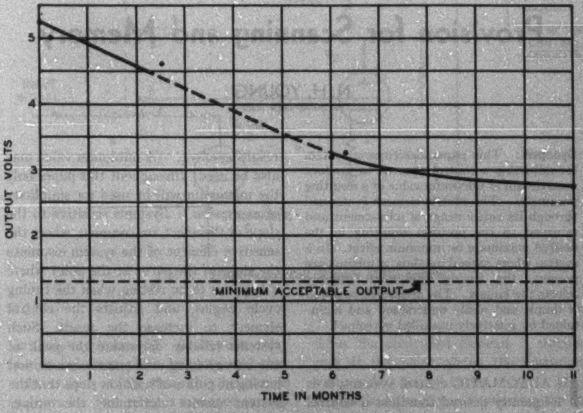


Fig. 10. Average output voltage of 96 phototransistors

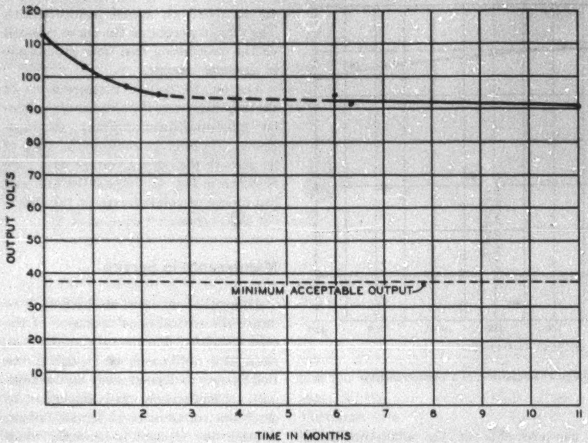


Fig. 11. Average output voltage of 96 phototransistor-amplifier combinations

An examination by the testman of this record will disclose any failure under this test condition.

Although this marginal test will disclose any channel approaching end of life gradually, there always remains the possibility of a sudden failure of some component during service operation. Such a failure will automatically cause a card record to be made of the conditions existing at the time of failure. If this record is caused by maloperation of a channel circuit, it will disclose which one failed thus permitting prompt corrective action to be taken. In the event this trouble occurred on a service call, the system is so arranged that the call will be completed using alternate equipment.

To determine which element has failed when an unsatisfactory channel is detected, the output voltages of the phototransistor and the amplifier are checked against the minimum values permitted and the weak unit can thus be isolated.

Conclusion

The test results show that the transistors and their circuits as used in the card translator are reliable. In addition, their service life appears to be satisfactory. The combination of these two characteristics is a vital factor, especially in view of the key function that translators perform in the *4A* toll switching system. The confidence with which this circuit is regarded is indicated by the fact that over one hundred translators will have been shipped to jobs by June 1, 1953. These installations represent the first commercial application of transistors in the Bell System.

References

1. Crossbar Toll Switching Systems, **L. G. Abraham, A. J. Busch, F. F. Shipley.** *AIEE Transactions (Electrical Engineering)*, vol. 63, June 1944, pp. 302–09.

2. Automatic Toll Switching Systems, **F. F. Shipley.** *AIEE Transactions*, vol. 71, pt. I, Sept. 1952, pp. 261–69.

No Discussion

An Automatic Control System with Provision for Scanning and Memory

N. H. YOUNG
NONMEMBER AIEE

Synopsis: This paper describes a system for adjusting a control element to produce a maximum or minimum value of a resulting parameter. The control element is scanned through its entire range of adjustment and returned to the position resulting in the desired maximum or minimum effect. In a system where several maxima or minima are present, the tuning equipment reliably selects the largest. The apparatus required is simple and easily understood and maintained by relatively unskilled personnel.

IN AUTOMATIC control systems, it is frequently desired to adjust a control element to produce a maximum value of a resulting effect. (A minimum value may also be used; throughout this paper only the maximum will be used for simplicity of expression.) Systems sensitive to the slope of the effect are common, where the sensitive element of the system examines the slope of the curve at the point where it happens to be resting when the tuning cycle begins, and adjusts the control element to increase the result. Such systems reliably determine the peak of any curve having only one maximum, and having no points of such low slope that the system cannot determine the proper direction of motion to correct.

In the system described in this paper, these limitations are not present. The system will accommodate response curves having more than one peak; in such cases the system will adjust the control element to the highest peak within the range. In addition, the system will operate correctly with curves having long, relatively flat sections within which the slope, if any, is so small as to be indeterminate. Curves of this type are often found in tuning radio transmitters and receivers. Secondary responses are occasionally produced by harmonics or image responses, and in receivers the system may be used to select

---

Paper 53-202, recommended by the AIEE Special Communications Applications Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted October 20, 1952; made available for printing March 30, 1953.

N. H. Young is with the Federal Telecommunication Laboratories, Inc., Nutley, N. J.

This paper is based on development work sponsored by the Bureau of Ships, U. S. Navy Department.

the strongest of several competing signals, where desired.

The sequence of operation is exactly like the operations performed by a careful human operator in tuning an equipment having the type of response described. Such an operator, when asked to readjust the system to tune to a new response point, would adjust the control element throughout its entire range, noting the amplitude of any peaks observed, then returning the control to the position producing the highest response. The automatic system proposed, which is called the "memory-scanning" system, follows exactly this procedure.

The elements of this system are shown in Fig. 1. The postulated response curve of the controlled device is shown as the solid line in Fig. 2. The input to the sensitive element of the control system is assumed to be a unidirectional voltage of varying magnitude. When a tuning cycle is initiated, the controlled element is moved to one end of its range (say, position 100 in Fig. 2). During this time the switch S in the system shown in Fig. 3 is in position no. 1, shorting the memory capacitor C. When the controlled element has reached the end of its range, a limit switch is operated, causing the switch S to go to position no. 2, and causing the motor to move the controlled element through its entire range (say, to position zero in Fig. 2). During this scan, the signal representing the response of the controlled element is being impressed through a diode D on the memory capacitor C. The potential on this condenser will rise when the input signal rises, but will not be able to drop when the input signal drops because of the unidirectional conducting properties of the diode D. Thus, the potential on the memory condenser will follow the dotted lines in Fig. 2, and at the end of this memory scan, will be equal to the highest potential the input signal had reached during the scan.

At the end of this scan, another limit switch operates, causing the switch S to be advanced to position no. 3. In this position, the memory capacitor is connected to one of a balanced pair of tubes, while the instantaneous input signal from the controlled element is applied to the other tube of the pair. The motor is also caused to begin to return the controlled element to the other end of its range. At the instant that the potential delivered by the controlled element equals the remembered potential stored in the memory capacitor C, the signals applied to the two tubes will be equal, and their plate potentials will be equal. (Necessary adjustments to affect such balance initially are assumed). A sensitive relay connected from plate to plate will thus release just at the time this balance is achieved. This relay is arranged to stop the motor moving the controlled element. Thus the controlled element is stopped just at the point giving the maximum value of the desired effect.

Practical design points affecting the accuracy of the foregoing statement are

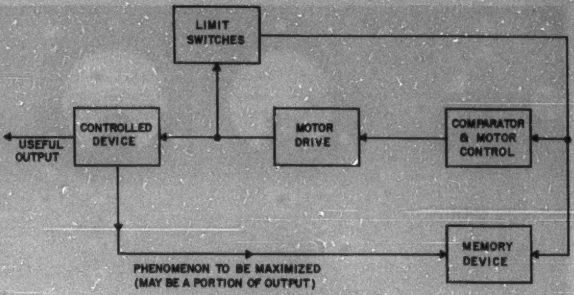

LIMIT SWITCHES
USEFUL OUTPUT
CONTROLLED DEVICE
MOTOR DRIVE
COMPARATOR & MOTOR CONTROL
MEMORY DEVICE
PHENOMENON TO BE MAXIMIZED
(MAY BE A PORTION OF OUTPUT)

Fig. 1. Block diagram of the system

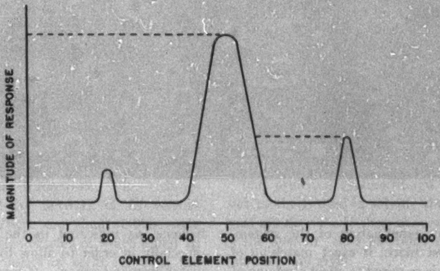


Fig. 2. Assumed response curve of a typical controlled element

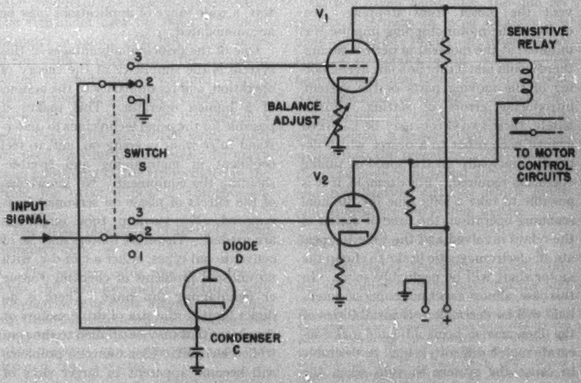


Fig. 3. Simplified circuit diagram of the system

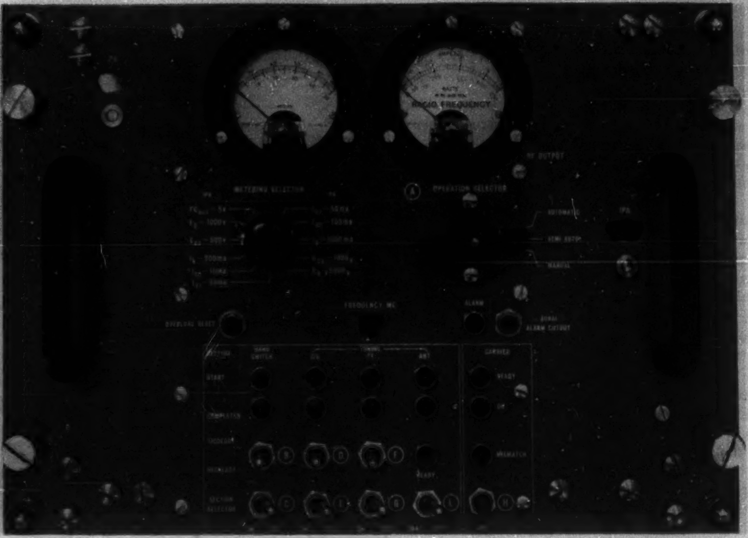


Fig. 4. Front panel of radio-frequency amplifier of transmitter embodying the system

evident by examination of the system. In the first place, if exact matching of potentials is required to effect stopping of the motor, even a minor change in sensitivity of the system between the memory scan and the final tuning scan might prevent the motor from stopping. Secondly, if the motor-stopping impulse was initiated at the moment of perfect tuning, there would inevitably be some overshooting of the moving parts of the system, introducing errors of setting. Fortunately, these two effects may be balanced against each other to a degree which depends principally on the speed of tuning operation required. For example, if it is possible to take a long time for the final scanning operation, the time of action of the relays involved and the time to operate an electromagnetic brake to clamp the motor shaft will be negligibly short. In this case, almost exact matching of potentials will be desirable. If instabilities of the phenomenon being adjusted make accurate match difficult, it may be desirable to cause the system to stop when the instantaneous voltage is equal to only 0.8 of the remembered voltage, and possibly introduce an intentional time lag in the application of the braking force, so that the normal stopping position of the system will be very near the peak. In other cases, where the time available for the tuning cycle is very limited, it may be necessary to anticipate the point of balance, as indicated in the foregoing, especially in order to allow time for the motor-stopping action. If necessary, the motor speed may be reduced for the final scan to increase the accuracy of adjustment. By suitable choice of design factors, a wide range of applications may be accommodated.

One of the principal advantages of this system is the simplicity of the theory of operation, and its similarity to the actions of a human operator. This makes it possible for the usual technicians to understand fully what is going on, and to feel confident in maintaining and trouble-shooting the equipment. No knowledge of the effects of phase on servomotors is required. No vacuum tube amplifiers are needed. The drive motors may be of conventional types, either a-c or d-c, with no unusual problems of checking torque or maintaining slip rings. There is no direct limit to the size of drive motors or the power that may be applied to the controlled element. (Some inertia problems will become apparent in larger sizes of apparatus.)

One embodiment of this principle that has reached production is the automatic tuning of the amplifier stages of a radio transmitter after the exciting frequency has been determined. One important advantage of this system over older systems having mechanical memory of the positions of shafts is in the indefinitely large number of predetermined channels that may be tuned. Any input frequency may be used, without regard to whether it has ever been set up before. Once the input is provided, the automatic system will resonate each of the amplifier stages to the proper frequency quickly and accurately.

The front panel of the radio-frequency amplifier section of this transmitter is shown in Fig. 4. It will be noticed that there are no tuning controls in the usual sense. The knobs shown are for switching the panel meter to the proper circuits, and for limiting the fully automatic action of the tuning system to slow down its tuning cycle for simpler analysis of troubles and for corrective maintenance. Jog switches are also provided to tune the circuits if the automatic system should fail, but they are not normally used. Two separate cycles of the above automatic tuning procedure are employed in this application. One motor drives the two low-level amplifier stages, which are ganged to a common shaft, while another motor drives the output amplifier, which supplies approximately 500 watts to a special automatic antenna resonating system. (The final amplifier cannot be ganged to the low-level stages because of the reaction of the antenna system on its resonance point.) The tuning cycle for the smaller unit, driving the low-level stages, requires approximately 6 seconds, while that of the higher-powered stage requires from 8 to 14 seconds, depending on the frequency band in which the equipment is being operated.

The system described leads to simple, easily maintained equipment, capable of setting any adjustable control to the point in its range producing a maximum (or minimum) of the effect desired. It is capable of an indefinitely large number of settings, depending on the input conditions. If the phenomenon being adjusted is susceptible to drift with time, repeating the tuning cycle will readjust to the new correct position. Its low cost and applicability to a wide range of effects makes it a valuable tool in the automatic adjustment of complex equipment.

No Discussion

The L3 Coaxial System

C. H. ELMENDORF
NONMEMBER AIEE

R. D. EHRBAR
NONMEMBER AIEE

R. H. KLIE
NONMEMBER AIEE

A. J. GROSSMAN
NONMEMBER AIEE

Synopsis: The *L3* coaxial system is a new broad-band facility for use with existing and new coaxial cables. It makes possible the transmission of 1,860 telephone channels or 600 telephone channels and a television channel in each direction on a pair of coaxial tubes. The principal system design problems and the methods used in their solution are described. The over-all system is described in terms of its components and their location in the system.

THE *L3* coaxial carrier system is a new broad-band transmission system capable of transmitting either 1,860 telephone message channels or 600 message channels and a 4.2-megacycle broadcast television channel, in each direction, on a pair of coaxials. The system is designed so that signals transmitted over any of these channels will meet high quality Bell System objectives after 4,000 miles of transmission.

The system is composed of auxiliary or line repeaters spaced at approximately 4-mile intervals along the cable route and connecting terminal or dropping repeaters where telephone or television signals are translated to or from the *L3* frequency band. Equalization equipment, power generating and power transmission equipment, and maintenance equipment are required at 100- to 200-mile intervals.

Planning and exploratory development for the system was started late in 1945 with the objective of designing a trunk route system which would provide the maximum channel capacity on the existing coaxial cable consistent with the state of the repeater art. At that time and for the next 4 years a large amount of new cable employing the 600-channel 3-megacycle *L1* coaxial carrier system was being installed or projected.[1]

Since a major field for the use of the *L3* system was to replace the *L1* system on existing routes, the design of the *L1* system, the cable, and the cable route layouts presented the *L3* system with a definite plant framework. Fig. 1 shows the present-day network of *L1* coaxial systems. There are about 8,000 route miles of cable installed of which about 70 per cent consists of eight coaxials, the remainder consisting of six and four coaxials. About 70 per cent of this cable uses coaxials with a 3/8-inch-diameter outer conductor, the present-day standard. The remainder uses the older 0.27-inch-diameter coaxials. All but a few miles of this cable is plowed into the ground or placed in underground conduit. Fig. 2 shows a piece of a typical 8-coaxial cable. Normally, the coaxials are included in a lead sheath with interstitial pairs, which are used for control purposes. In many cases additional quads are included in the cable for other types of transmission systems.

The broad objectives of the *L3* system planning were:

1. The existing cable was to be reused. Thus, the cable loss and its variation with temperature, the cable irregularities resulting from manufacturing and splicing, and the power transmission capabilities of the cable became basic restrictions on the design of the *L3* system.
2. The *L1* telephone terminal equipment was to be reused. This equipment involves channel banks, group and supergroup equipment, and carrier supplies.[3] This limited the system planning to the use of frequency division multiplex on a single side-band carrier suppressed basis.
3. It would be desirable to reuse existing *L1* repeater locations and buildings. The *L1* auxiliary repeaters are spaced at 8-mile intervals and housed in 6-foot by 9-foot concrete block huts. The *L1* main repeaters are spaced at 40-to-160-mile intervals largely dictated by geographical and power transmission considerations.
4. Sufficient bandwidth should be provided so that a black-and-white television signal of at least 4-megacycle quality could be transmitted simultaneously with 600 message signals, the message capacity of the *L1* system. Alternatively, as many massage channels as possible should be transmitted when there is no need for television service.
5. The channels should meet Bell System high quality signal-to-noise and equalization objectives after 4,000 miles of transmission.

The principal system design problems and methods used in solving these problems are described; then the components of the system, their locations, and their functions are given.

Transmission Design

With a given cable loss, the line repeaters determine in large measure the bandwidth and quality of transmission and the economics of the system. The basic system plan therefore evolves from a consideration of the signal-to-noise and equalization performance—that is, the transmission stability—that can be designed into the repeaters. This leads to the development of broad signal-to-noise and equalization analyses which guide and co-ordinate the system design.

SIGNAL-TO-NOISE DESIGN

Simply stated, the signal-to-noise problem is to adjust the repeater spacing and

---

Paper 53-221, recommended by the AIEE Wire Communications Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 18, 1953; made available for printing April 27, 1953.

C. H. ELMENDORF, R. D. EHRBAR, R. H. KLIE, and A. J. GROSSMAN are all with the Bell Telephone Laboratories, Inc., Murray Hill, N. J.

Many Bell Laboratories departments co-operated in developing the complex *L3* system. In a period of 7 years, over 100 engineers and technicians contributed to its design. In the system planning aspects covered in this paper, particularly large contributions were made by L. G. Abraham, C. H. Bidwell, and S. E. Miller. The section on "Automatic Switching" was written by P. T. Sproul, and the section on "Terminals" by C. G. Arnold.

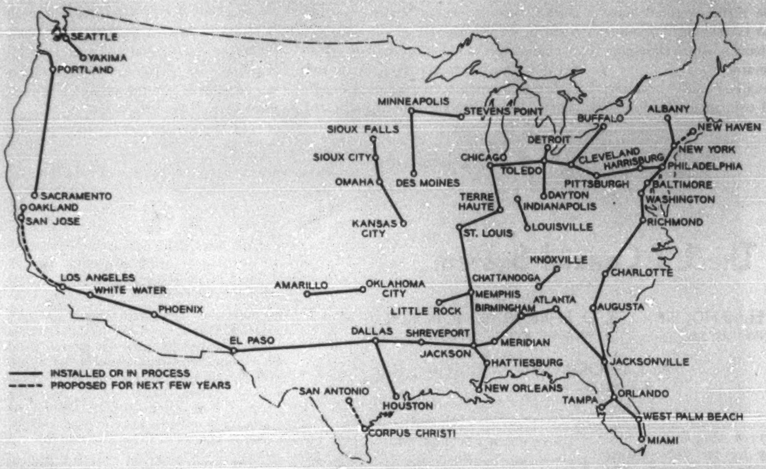

SEATTLE
YAKIMA
PORTLAND
SACRAMENTO
OAKLAND
SAN JOSE
LOS ANGELES
WHITE WATER
PHOENIX
EL PASO
MINNEAPOLIS
STEVENS POINT
SIOUX FALLS
SIOUX CITY
OMAHA
DES MOINES
KANSAS CITY
AMARILLO
OKLAHOMA CITY
DALLAS
SHREVEPORT
JACKSON
SAN ANTONIO
HOUSTON
CORPUS CHRISTI
CHICAGO
TOLEDO
DETROIT
BUFFALO
CLEVELAND
HARRISBURG
PITTSBURGH
DAYTON
TERRE HAUTE
INDIANAPOLIS
LOUISVILLE
ST. LOUIS
KNOXVILLE
CHATTANOOGA
MEMPHIS
LITTLE ROCK
BIRMINGHAM
ATLANTA
MERIDIAN
HATTIESBURG
NEW ORLEANS
ALBANY
NEW HAVEN
NEW YORK
PHILADELPHIA
BALTIMORE
WASHINGTON
RICHMOND
CHARLOTTE
AUGUSTA
JACKSONVILLE
ORLANDO
TAMPA
WEST PALM BEACH
MIAMI
INSTALLED OR IN PROCESS
PROPOSED FOR NEXT FEW YEARS

Fig. 1. Coaxial cable routes

bandwidth of the system so that channel objectives can be met with the repeater noise, linearity, and gain performance that the electron tube and circuit art permit. In detail this means the following:

1. translating the broad transmission objectives on message and television channels into detailed requirements on noise, specific modulation products, and compression;
2. analyzing the amount of these interferences that result from various repeater design choices;
3. determining the effect of signal wave form and frequency allocations on both the channel requirements and the repeater performance; and
4. integrating these studies into a specific system design plan that meets the objectives.

Telephone Channel Interference Objectives

The amount of noise, tone interferences, or crosstalk that is considered tolerable in telephone channels is generally determined by judgments involving the subjective reactions of representative observers to specific interferences on typical transmitted signals, and by the cost of providing a given grade of service. The broad objectives for message channels stem from early unpublished work on transmission standards. The interference and load capacity requirements for transmission systems involving large numbers of message channels were developed by Dixon, Holbrook and Bennett.[3,4] In effect, they provide techniques for translating channel objectives into linearity and power-handling requirements on repeaters, taking into account the statistical properties of individual and multichannel speech. Based on the data and techniques in these articles, the requirements on individual channels shown in Table I can be derived. These requirements in themselves form an important basis for the signal-to-noise design of the system. However, in a highly refined system design, it is necessary to extend our notions of requirements somewhat further.

In the *L3* signal-to-noise design the message channel requirements of Table I were used as the initial basis for study. However, when specific interferences of a complex nature were found to be limiting, the wave forms and the probability of their occurrence were examined in detail. As a result of these studies, two distinctive types of interferences were found to be important when the system is used to carry message and television signals simultaneously. The first of these, due to both second- and third-order modulation involving multifrequency key pulse signals and components of the television signal, has the characteristics of intermittent musical tones. The second, due to the second-order difference products generated by the television signal components, produces tones in the message channels which vary in amplitude and frequency as the television signal changes with picture content. Both types of interference were generated in the laboratory and recorded on tapes. From these tapes, records were cut and then used in a series of subjective tests which were made to determine the maximum permissible magnitudes consistent with other important message circuit objectives.

Television Channel Interference Objectives

The amount of noise and single-frequency interference that can be tolerated in a commercial grade of television channel again depends on judgments involving the subjective reactions of observers and the cost of providing a given grade of service. The broad objectives are based on subjective measurements by Mertz and Baldwin.[6–7] From this work it has been determined that 95 per cent of the observers consider a signal-to-noise ratio of 40 decibels (composite signal to rms noise) tolerable, providing the noise has a frequency characteristic that rises about 11 decibels across the video band. Likewise the tolerable single-frequency interference can be set at −70-decibel (peak sine wave below composite signal) if the interference falls below about 1 megacycle. However, the requirement becomes more lenient for interferences falling in the upper part of the band.

Again, for a refined system design, more detailed account must also be taken of the requirements on short duration interferences, the probability of interference occurring, and the exact frequency in the television spectrum at which an interference occurs.

In the *L3* signal-to-noise design the broad television channel objectives outlined in the preceding were used except when a specific complex interference was

Table I. Summary of Message Circuit Objectives

Allowable Zero Level Interference in 3-Kc Band

Source of Interference	Type of Interference	Dba, Message Weighting	Decibels below 1 Milliwatt,* Unweighted
Terminals	Largely spillover between channels, cross-modulation, and crosstalk	+32	50†
Line	Noise and multichannel modulation	+36	46†
Line	Unintelligible crosstalk and babble	+24	58†
Line	Tones	+24	61‡
Total	All sources	+38	44†

* The translations from dba to decibels below 1 milliwatt are effected by noting that a 3,000-cycle band of flat noise with 1 milliwatt of power equals +82 dba and that 1 milliwatt of 1,000-cycle single frequency is equal to +85 dba.

† Interference assumed evenly distributed over 3,000-cycle band.

‡ Tones assumed to be at 1,000 cycles.

found to be limiting. For complex interferences, three additional types of requirement data were used: 1. Tests were made to determine visual thresholds relative to steady tones of short bursts of energy such as occur in the television channel as a result of switchhook "bang-up"; and multifrequency key-pulsing signals in the message channels. Fig. 3 shows the relation between the steady-state and transient requirement. 2. Advantage was taken from the fact that interferences falling between the 15.75-kc line scan multiples of the television signal would be less interfering than unwanted energy falling directly at the line scan multiples. 3. A judgment was used that the tolerability of an interference depends on its probability of occurrence. The judgment was not made on a quantitative basis but when an interference was found to exceed its requirement by a few decibels two or three times a day it was ignored in the signal-to-noise design.

Frequency Allocations

The final frequency allocations shown in Fig. 4 are a result of the signal-to-noise design. The principal features were determined on rather general grounds. When the system is arranged for combined television-message transmission, the television channel is placed above the message channels so that the second harmonic of the television carrier and its immediately adjacent side bands will fall at the top edge of the band where the requirement is more lenient. Likewise the line repeater noise tends to rise with frequency as does the amount of noise that the television channel can tolerate. Details of the frequency allocations shown in Fig. 4 will be discussed later.

Pilot frequencies, indicated in Fig. 4, are transmitted to control the transmission characteristic of the system as described in a companion paper.[8] The frequencies, and the power at which the tones are transmitted, were selected on two bases:

1. Where possible, frequencies used for similar purposes in the *L1* system were selected for possible economies in pilot supply design and manufacture; these are the 556-, 2,064- and 3,096-kc pilots.
2. The transmission of these pilots should not materially degrade the signal-to-noise or load capacity performance of the system.

The latter requirement led to a careful study of cross-modulation products involving the pilot frequencies to assure that message and television objectives would be met.

Repeater Performance

The details of the amplifier design and the factors which determine its performance are covered in a companion paper.[9] For purposes of the signal-to-noise design it is sufficient to know the noise power versus frequency characteristic, the second- and third-order modulation coefficients of the repeater as functions of frequency, and the overload performance of the repeaters. These factors depend on the repeater spacing and cable loss characteristic, electron tube parameters, achievable feedback, and the bandwidth to be transmitted. Thus, in the design procedures the dependence of these properties on repeater gain and bandwidth are determined and used in adjusting the system parameters for a final compatible design. Figs. 5 and 6 show the noise and linearity properties of the final *L3* repeater. The 4-mile repeater spacing requires a repeater gain shown on Fig. 7.

Signal Mechanisms

A signal-to-noise plan which contemplates transmitting the complex wave form of the combined telephone and television channels through 1,000 auxiliary amplifiers and about 200 flat amplifiers with performance factors that are variable with frequency will depend very strongly on the detailed analysis of the interactions between the signals and the repeater system characteristics. In developing this aspect of the signal-to-noise design four related phenomena had to be examined in detail.

Intermodulation Between Signals in Different Parts of Band. In the classical multichannel modulation theory for a large number of message channels, the modulation noise generated by interaction of the speech signals owing to the nonlinear characteristics of the amplifier is shown to be equivalent in interfering effect to random noise. In addition to this type of interference in message channels, cross-modulation between components of the message and television signals result in a host of specific individual modulation products which have been examined by determining their amplitude, duration, and probability of occurrence, and then relating them to the requirements previously discussed. Approximately 400 different products or groups of products were studied in the design of the *L3* system. All but about thirty of these were found to be of negligible importance for the signal levels and frequency allocations being given serious consideration. On final analysis six of these thirty products were

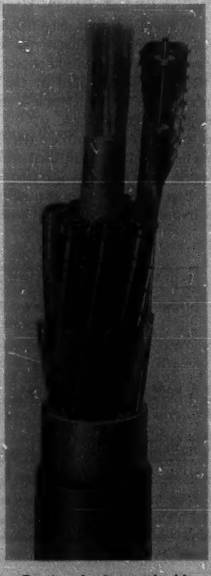


Fig. 2. An 8-coaxial cable

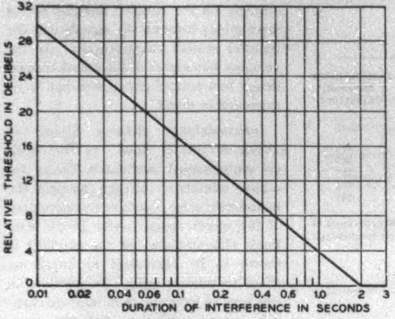


Fig. 3. **Television bar pattern threshold versus duration**

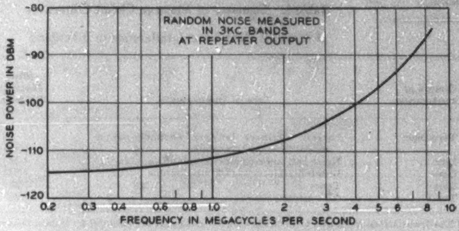


Fig. 5. **L3 auxiliary repeater noise characteristic**

found to be controlling in establishing system levels. Fig. 8 shows the generating signals and the products they form for the six most critical products. The exact way in which the critical products entered into the determination of signal levels and frequency allocation will be discussed later.

Location of Television Carrier Relative to Telephone Channel Carriers. Among the important modulation product types is one formed by difference frequencies involving components of the telephone and television signals, see Fig. 8(D). These interferences fall back into the telephone band and are of different magnitudes depending, among other things, on which components of the television signal produce them; those produced by the television carrier and adjacent line scan multiples are by far the strongest.

The energy in a disturbing telephone channel tends to be concentrated near the 1,000-cycle point in the voice frequency band. By careful choice of the television carrier frequency, the difference products produced by cross-modulation between telephone signals and the high-magnitude television signal components can be made to fall at frequencies such that the high-energy portions of these products are greatly attenuated by the cutoff characteristics of channel filters.

The message channels are spaced at 4-kc intervals controlled by carrier frequencies which are multiples of 4 kc. To obtain the maximum advantage from the channel filter cutoff characteristic as described, it was found desirable to set the television carrier frequency 1 kc below a 4-kc multiple. A direct result of this allocation is a gain of 12 decibels in television signal-to-noise performance over what could be realized if the carrier had been set at a 4-kc multiple. Such an allocation would have required a 12-decibel lower magnitude of television signal in order to meet the message channel objectives.

Addition of Modulation Products Along the Line. It has been established by analysis and experiment, that in a multirepeater system second-order modulation products tend to accumulate on a power basis while certain third-order products tend to add on a direct or voltage basis. This direct addition of third-order products depends on the slope of the phase curve being the same over small frequency intervals from repeater to repeater. In multichannel telephone systems, the locations of channels in the frequency band are shifted at intervals along the line to avoid this direct addition of third-order products. In the combined telephone-television application of the *L3* system the $A+B-C$ product illustrated in Fig. 8(G) is formed. Since the B and C components are television line scan multiples which cannot be shifted in location, certain components of this type of product would add directly in a 4,000-mile system. If this were allowed to take place, the requirements would be exceeded by many decibels. However, by placing the delay distortion equalization only in the television band at approximately 200-mile intervals, the phase of these products can be shifted so that rms addition of products accumulated over several 200-mile links of the system may be assumed.

Wave Form of Transmitted Television Signal. Early studies of *L3* led to the

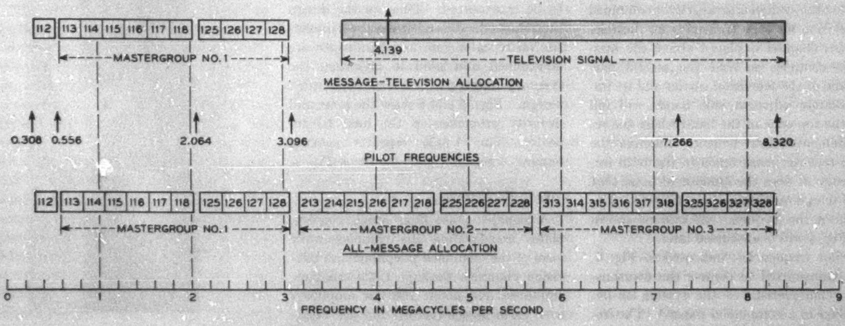

112
113
114
115
116
117
118
125
126
127
128
MASTERGROUP NO. 1
4.139
TELEVISION SIGNAL
MESSAGE-TELEVISION ALLOCATION
0.308
0.556
2.064
3.096
7.266
8.320
PILOT FREQUENCIES
112
113
114
115
116
117
118
125
126
127
128
213
214
215
216
217
218
225
226
227
228
313
314
315
316
317
318
325
326
327
328
MASTERGROUP NO. 1
MASTERGROUP NO. 2
MASTERGROUP NO. 3
ALL-MESSAGE ALLOCATION
0
1
2
3
4
5
6
7
8
9
FREQUENCY IN MEGACYCLES PER SECOND

Fig. 4. **L3 frequency allocations**

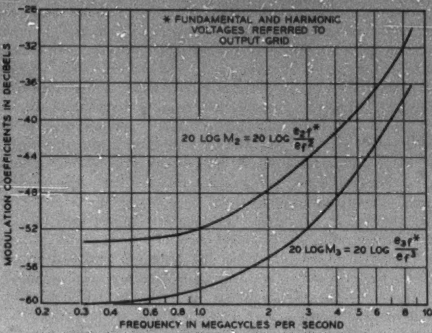


Fig. 6. L3 amplifier modulation coefficients

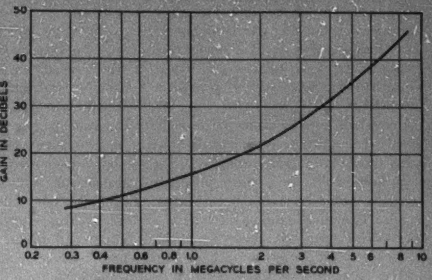


Fig. 7. L3 repeater gain characteristic

conclusion that the most economical method of transmitting the television signal would be by amplitude modulation of a carrier with one side band partially suppressed, that is, vestigial side-band transmission. There remained, however, three major problems for detailed study:

1. the transmission of d-c components of the video signal;

2. the per-cent modulation of the carrier which for convenience is defined in terms of excess carrier ratio, the ratio of the peak (white) signal to the peak-to-peak composite signal as measured in the carrier frequency envelope; and

3. the sign or sense of modulation, that is, whether increasing or decreasing brightness should correspond to increasing signal voltage on the high-frequency line.

Typical wave forms illustrating the alternatives are given in Fig. 9.

The solution to each of these problems required an understanding of how the various alternatives would be affected by the system noise and linearity performance and an understanding of representative television-viewing tube performance with respect to susceptibility to different types of interference. In analyzing the effect of system performance on these problems, it was found that nonlinearity (cross modulation) would produce interferences in the television band which, while they are very complex electrically because of the effect of cross modulation involving line scan components of the signal, would produce the same effect on viewing tubes as single-frequency interferences, that is, bar patterns. Further simplifications were made in the analysis when it was found that such interferences were most visible in relatively large areas of television pictures having essentially constant brightness. During the time intervals corresponding to such areas, the video frequency voltage of the television signal is essentially constant and therefore, in the cases of interest, it could be assumed that the magnitude of the television carrier would also be constant during such intervals. Thus, to compute the magnitude of any modulation product which falls into the television band and which has as one of its components the television signal itself, it is found convenient to use in the computation the magnitude of the television carrier corresponding to either black or white portions of a picture signal. (The reason for intermediate shades of gray being less susceptible than either black or white will be discussed.)

To evaluate the effect of television-viewing tubes on wave form problems, a number of tests were made to determine blank field threshold values of single-frequency interference as a function of frequency for typical viewing tubes. Furthermore, judgments were made as to what might be expected of future viewing tubes with respect to achievable high light brightness, contrast ratio, and operating characteristics. As a result of these tests and judgments, a series of requirements were derived on the basis of long-range objectives to be met for these projected characteristics. The results of these tests and judgments are summarized as follows:

Television viewing tube characteristics assumed for *L3* signal-to-noise analyses.

1. Brightness-grid voltage characteristic of viewing tubes follows 5/2 power law

$B \alpha e_g^{5/2}$

2. Maximum high light brightness of viewing tubes will be 150-foot lamberts.

3. Contrast ratio of viewing tubes will be 150 to 1.

4. Viewing tubes will have interference sensitivities which vary with brightness in accordance with the characteristic of Fig. 10.

5. The visibility of bar patterns will decrease with frequency in accordance with the characteristic of Fig. 11.

6. Deviations from the Weber-Fechner law may be assumed to follow the curve of Fig. 12. This law states that "the minimum change in stimulus necessary to produce a perceptible change in response is proportional to the stimulus already existing."

Using the parameters and methods of analysis outlined in the preceding paragraphs, the relative system performance achievable with each of the carrier frequency wave forms of Fig. 9 was computed or determined by observation. For example, these wave forms are all drawn to the same peak-to-peak amplitude. If we assume that the coaxial system is limited only by the peak amplitude transmitted, we may use Fig. 9 to determine relative signal-to-noise performance directly by measuring the peak-to-peak magnitude of the composite signal voltage (sync tip to white) transmitted.

Fig. 9 may also be used to obtain relative modulation performance. For this purpose, the following factors must be considered:

1. the magnitude of the signal generating the interference ("black" or "white" carrier magnitude);

2. whether the interfereuce is proportional directly or to the square of the carrier magnitude;

3. relative interference sensitivity in black or white portions of the picture; and

4. deviations from the Weber-Fechner law as the brightness is varied over its full range.

The relationships among these factors were used to establish that for all cases of interest, bar patterns caused by cross-modulation are always more interfering in either black or white portions of a picture than in an intermediate gray area.

Table II shows the relative system performance for the five carrier frequency wave forms of Fig. 9. For comparison purposes, the signal-to-noise and signal-to-bar pattern ratios are all related to Fig. 9(F).

It is obvious from Table II that the signal is transmitted most efficiently at an excess carrier ratio (ECR) of one half. The wave form of Fig. 9(F), which illustrates excess carrier of one-half, is the one used in *L3*. Television terminal circuit problems arising from this choice of carrier frequency wave form are discussed in another paper.[10]

Signal Levels and Repeater Spacing

In a broad-band system like *L3*, the problem of determining the repeater spacing is made complex by the large number of parameters that must be considered. The approach to this problem that has been used to advantage in the *L3* design is to assume several reasonable values of repeater spacing and determine for each the system performance achievable with various combinations of important parameter values. This method also permits evaluation of the effects on repeater spacing caused by variations in parameters so that it is possible to form judgments as to the most economic design.

One of the important factors in setting repeater spacing is the magnitudes at which signals are transmitted in the system and the relation between these magnitudes and signal-to-noise and repeater overload performance. In the all-telephone system (1,860 channels), the telephone levels (decibels with respect to the transmitting toll test board) were set to optimize signal-to-noise performance. To avoid penalizing the channels in the upper part of the band, where random noise tends to be much higher than at low frequencies, the levels of the three mastergroups are staggered. At the output of any repeater in the high-frequency line, the nominal level of mastergroup 1 is −21 decibels, that of mastergroup 2 is −16 decibels, and that of mastergroup 3 is −11 decibels. As a consequence of setting levels in this way, the random noise in the message channels is approximately 2 decibels higher on the average than modulation noise. It can be shown that with both second- and third-order modulation products contributing, and with third-order somewhat predominant, this relation between random noise and modulation noise produces optimum signal-to-noise performance. With these levels, and 1,860 channel telephone system has approximately a 6-decibel margin against repeater overload which, for *L3* purposes, has been defined as the point at which the repeater modulation coefficients just depart from their constant small-signal values. The signal-to-noise objective of +29 decibels adjusted at the −9-decibel level is met with about a 2-decibel margin.

When the system is used to transmit television and message signals simultaneously, the level of the telephone channels in mastergroup 1 at the repeater output is the same as that of mastergroup 1 in the all-telephone application, −21 decibels. The most convenient measure of the television signal is the power of the unmodulated carrier at the output of a repeater. Its value is 6 decibels above 1 milliwatt. Due to the interrelations between the two signals, the limitations on achievable maximum transmission levels or magnitudes arise from certain types of second-order modulation products

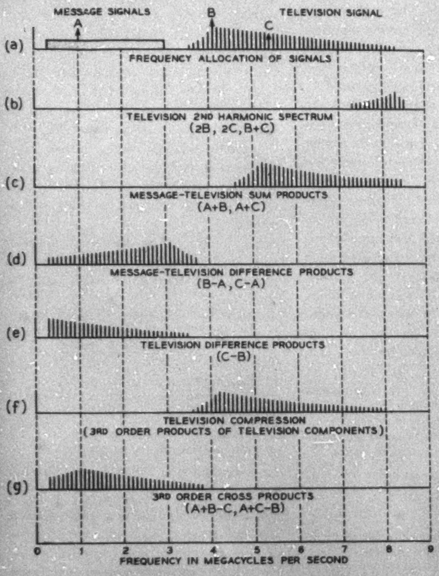


Fig. 8. L3 coaxial system—critical modulation products in combined-message-television application

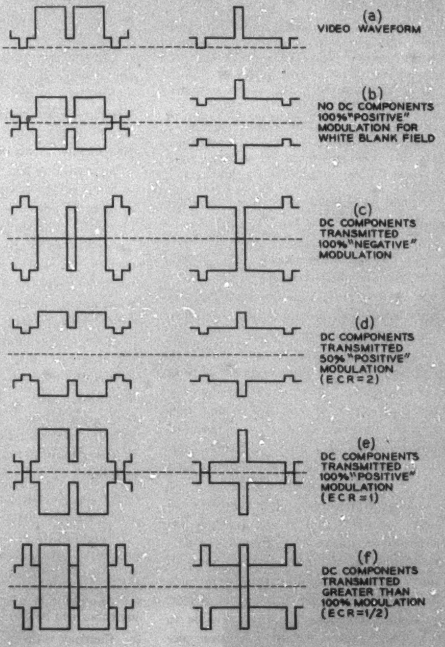

(a)
VIDEO WAVEFORM
(b)
NO DC COMPONENTS
100% "POSITIVE" MODULATION FOR WHITE BLANK FIELD
(c)
DC COMPONENTS TRANSMITTED
100% "NEGATIVE" MODULATION
(d)
DC COMPONENTS TRANSMITTED
50% "POSITIVE" MODULATION
(ECR=2)
(e)
DC COMPONENTS TRANSMITTED
100% "POSITIVE" MODULATION
(ECR=1)
(f)
DC COMPONENTS TRANSMITTED
GREATER THAN 100% MODULATION
(ECR=1/2)

Fig. 9. Typical television signals-alternative carrier frequency wave forms

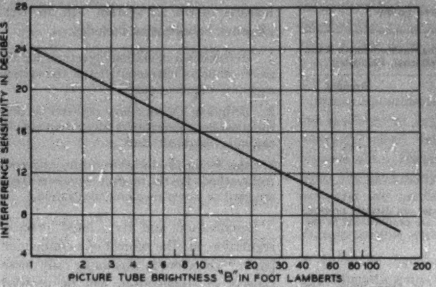


Fig. 10. Picture tube interference sensitivity assumed for L3

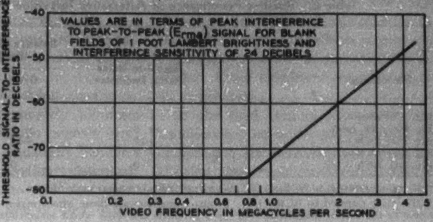


Fig. 11. Threshold values for bar patterns

rather than from optimizing signal-to-noise performance as in the all-telephone application. One of these types consists of sum products of cross-modulation between telephone and television signal components. These form bar patterns and, in so far as signal-to-interference ratio is concerned, are independent of the television signal magnitude. Thus, adjusting such products to equal the appropriate requirement has the effect of setting the maximum permissible magnitude or level of the telephone signal. The second type of limiting product is due to difference frequencies formed by cross modulation among the television signal components. These fall into the telephone channels and, after the telephone level has been set as described, permit calculation of the maximum permissible television signal magnitude.

With signals set at a −21-decibel level for telephone and 6 decibels above 1 milliwatt unmodulated carrier for television, all of the critical products discussed in the preceding section and illustrated in Fig. 8 have adequate margin. The 40-decibel signal-to-noise objective for 4,000-mile television transmission is met with about a 2-decibel margin, and long haul (4,000-mile) message channels meet the +29 decibels adjusted at the −9-decibel level objective with about a 5-decibel margin. A margin of about 5 decibels is also realized with respect to repeater overload performance.

The single-frequency pilots are adjusted to have the following values of power at the output of a transmitting amplifier:

7,266 kc	16 decibels below 1 milliwatt
8,320 kc	26 decibels below 1 milliwatt
all others	36 decibels below 1 milliwatt

With these values, modulation products produced by cross-modulation among the pilots, as well as message signals and television signals, meet the appropriate objectives.

Frogging of Message Circuits

When signals, either message or television, are transmitted over long distances through many amplifiers in tandem the accumulation of modulation products along the line becomes an important system problem for two reasons:

1. The accumulation of certain types of third order products tends to follow a direct or in-phase law.
2. The distribution of modulation products over the band produces more modulation noise in certain parts of the band than in others.

Both of these cumulation problems are alleviated if, at intervals along the line, the signals are shifted with respect to one another in the band, a process known as frogging.

In the *L3* system, signal-to-noise performance is substantially improved by frogging the supergroups at intervals of about 800 miles. In the 1,860 channel all-message application, the busy-hour signal-to-noise performance of 4,000-mile circuits is alike to within 2 decibels with all channels meeting the objective of +29 decibels adjusted at the −9-decibel level. In contrast, if frogging were not specified, a substantial number of circuits, or (10 to 20 per cent) would fail to meet the objectives, while the performance of other channels would be better than required by 6 decibels or more.

When the system is used for combined message-television signals, the message circuits are frogged in supergroup blocks at approximate 800-mile intervals except for supergroups 113 and 114, which must be frogged at 400-mile intervals. This procedure is necessary to prevent second-order sum and difference products of message and television signal components from cumulating excessively, especially those products which involve television signal components close to the television carrier. Frogging these supergroups more frequently than others results in a 3-decibel improvement in television signal-to-noise performance.

Special Services Transmission

During the early design stages, requirements based on the transmission of message and television signals were used to set repeater spacing, to determine the bandwidth and frequency allocations, and to fix important design parameters of the amplifiers. Concurrently, the objectives for the transmission of telegraph, program, and telephotograph signals were studied, and, before the system design crystallized, analyses were made to assure that these special services objectives would be met.

In a few instances it was found that the special services objectives tended to dominate and the system requirements and design were adjusted accordingly. For the most part, however, channels which meet message circuit objectives are satisfactory for special services transmission. In *L3*, telegraph and telephotograph signals may be transmitted without restriction provided the proportion of

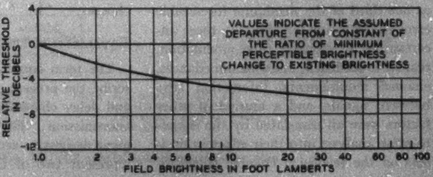


Fig. 12. Assumed deviation from Weber-Techner law

Table II. Relative Performance of Alternative Television Wave Forms

Wave Form	Relative Signal-to-Noise Ratio, Decibels*	Relative Signal-to-Modulation Ratio (Bar Patterns), Decibels†	
		Group 1	Group 2
Fig. 9(b) no d-c	+10.2	+11	+11.3
Fig. 9(c) neg. mod.	+ 6	+ 9.5	+12.5
Fig. 9(d) ECR = 2	+12	+14	+15.5
Fig. 9(e) ECR = 1	+ 6	0	+ 6
Fig. 9(f) ECR = 1/2	0	0	0

* All values referred to ECR = 1/2; plus values indicate poorer performance.

† All values referred to ECR = 1/2. Group 1 products are those whose magnitudes are directly proportional to the carrier magnitude. Group 2 products are those whose magnitudes are proportional to the square of the carrier magnitude.

these signals does not materially exceed the proportion now installed in the plant. Program signals may be transmitted in the 1,860 all-message arrangement without restriction but, when television transmission is provided, program circuits are restricted to supergroups 113 and 114. This restriction is due to the fact that program circuits are usually more than 4 kc wide; interferences of high magnitude which normally fall between 3,300 and 4,000 cycles or below 300 cycles in message channels of supergroups other than 113 and 114 would fall close to 4,000 cycles in a program circuit where there is high susceptibility to interference.

Uncertainties

In the early stages of system design, firm decisions have to be made on such matters as repeater spacing, bandwidth, and component characteristics. These decisions must be based on a detailed signal-to-noise analysis which in turn involves many judgments of repeater performance parameters, tolerable system requirements, and the effects of signal mechanisms on system performance. It would be easy and safe to engineer the system to provide enough signal-to-noise margin to cover the uncertainties in each of these judgments. Conservative engineering of this type could easily have justified a repeater spacing of 3 miles instead of the 4 miles actually chosen. Instead, an effort was made to estimate a mid-range, or most probable, value for each performance, requirement, or mechanism factor entering into the signal-to-noise design. In addition, a probable uncertainty was estimated for each critical parameter. This was usually taken as one-third of the maximum foreseeable error in the estimate. Finally, these uncertainties in electron tube modulation, realizable feedback, network impedances, channel requirements, interaction laws between signals, and a myriad of other factors were all translated by the signal-to-noise analysis into their effect in decibels on the television channel signal-to-noise performance. On this basis, the probable uncertainties were summed on an rss basis to find the probable uncertainty in the over-all design. Whereas the direct addition of the probable uncertainties gave a figure of about 20 decibels uncertainty in the design, the rss addition indicated about 6 decibels uncertainty. It was then argued that during the ensuing years of development the probability of finding all the judgements to be wrong in the same sense was extremely small. On the other hand, it was deemed reasonable to provide enough margin so that there would be perhaps a 75-per-cent chance of not exceeding the margin. Six decibels of margin was therefore provided, half by clear margin and half by having available economically feasible changes in system design such as a decrease in the telephone channel frogging interval. Any further error in judgment would then have to be taken up by degrading performance below the desired objectives. As the system design proceeded, the early judgments were changed in considerable measure. Likewise numerous additional system parameters were introduced. However, at no point in the system planning was the balance of factors such that there was less than 3 decibels clear margin.

Margin handled in this way becomes a carefully husbanded asset of the whole system. In designing or analyzing a part of the system, a major effort must be made to achieve the performance introduced into the initial determination of repeater spacing and bandwidth. The design of each individual part of the system cannot be allowed a margin which can be used up as the individual designer chooses.

EQUALIZATION DESIGN

The term equalization is used to describe the process of obtaining flat gain and delay characteristics for the system transmission. The system and equipment designs to accomplish this function represent two of the major engineering features of the *L3* system. In an over-all sense, equalization includes:

1. determining deviation objectives for the gain and delay characteristics of the system and its component parts;
2. designing the auxiliary repeater so that the most economical over-all system equalization is obtained; and
3. specifying the location, form, and control methods for the mop-up equalizers that are used at intervals along the system.

Equalization and its related process, regulation, are the subject of a companion paper;[8] therefore in this paper only those aspects of equalization will be covered which are necessary for an appreciation of the over-all system design.

Transmission Objectives

The requirements on the gain characteristic of a band used for multichannel telephony depend on two message channel objectives. One of these is that the gain of a message channel must not vary by more than 2 decibels over the 4-kc band. To meet this requirement, broad changes in the transmission characteristic of the message band are held to less than 0.5 decibel for 150-mile links. The second objective stems primarily from the need to transmit telephotograph signals. Since these signals are relatively intolerant of level changes, the transmission characteristics of working lines and protection lines are made alike to within ±0.25 decibel.

The requirements on the gain and delay characteristics of the television band are based on the subjective determination that an echo delayed by about 2 microseconds or more in a representative picture is considered tolerable by 95 per cent of the viewers when the peak-to-peak voltage of the echo signal is 39 decibels below the peak-to-peak signal voltage.[11] The translation of this echo objective to allowable variations in the gain and delay characteristics is straight forward if idealized sinusoidal deviations extending across the whole band are assumed.

In practice, the characteristics of the transmission deviations in a long repeatered system are very complex and therefore, the idealized objectives are only a tentative guide in system design. Since we do not have a thoroughly satisfactory method of evaluating complex echo patterns, the exact nature of the final television mop-up arrangements will be determined after subjective tests have been made on the interfering effects of echoes resulting from the complex transmission deviations of representative links of the system.

The Mop-up Plan

The deviations from ideally flat gain and delay transmission characteristics may be classified in three broad categories: 1. fixed deviations; 2. slowly varying deviations; and 3. rapidly varying deviations. The distinction made between slow and rapid in the last two categories relates to the frequency of adjustment needed to meet system objectives. Those variations which require adjustment more often than once a week are considered rapid, and those requiring adjustment at longer intervals are considered slow.

Corresponding to each of the three classifications of deviations is a set of equalizers, fixed, manually adjustable, or automatic under control of the pilot or a temperature sensitive element. Networks capable of fulfilling the functions of each are distributed along the line according to carefully prepared rules which enable system objectives to be economically met. The locations of these equalizers, their functions, and general characteristics, are illustrated in Fig. 13.

Fixed Equalizers. To the extent that the auxiliary repeater is designed so that its nominal gain compensates for the loss of 4 miles of coaxial, it may be considered as the first step of fixed equalization. In addition to the amplifier, the auxiliary repeaters are equipped with artificial lines, which are used to build out the loss of short sections to the equivalent of 4 miles of cable, and basic equalizers which provide for differences in the loss characteristics of different types of cable.

The second and final step of fixed gain equalization is known as a design deviation equalizer. Its function is to correct accumulated deviations caused by the failure of the average auxiliary repeater to compensate exactly for cable loss. These equalizers will be used at every mop-up point, at 40- to 120-mile intervals.

When television is transmitted, fixed delay equalizers are used at approximately 150-mile intervals. These equalizers compensate for the delay distortion introduced by the cutoffs of the auxiliary repeater sections.

Manually Adjustable Equalizers. The manually adjustable gain equalizers consist of networks whose loss-frequency characteristics are related to one another by a Fourier series type of representation. The number of terms of the series required to meet system objectives varies with different types of mop-up points, and depends on the system application being provided for, all-message or combined message-television service. Equalizers of this type are used in mop-up points at 40- to 120-mile intervals along the line.

Manually adjustable delay equalizers are provided at approximately 150-mile intervals when television signals are transmitted. These equalizers supplement the fixed delay equalizers described here and are needed to trim the delay characteristic of the line in finer detail than is possible with fixed equalizers.

Automatic Equalizers. The first step of automatic-gain equalization is provided at each auxiliary repeater. The nominal gain characteristic of the repeater is designed to match the loss characteristic of the coaxial at 55 degrees Fahrenheit. The cable loss varies with changes in temperature; the variations, however, have a predictable characteristic, being very closely proportional in decibels to the square root of frequency. To compensate for these changes, the gain characteristic of the amplifier at auxiliary repeaters is adjustable and follows the loss of the cable under the control of a thermistor. Two types of circuits are used to control the current fed to the thermistor as described in a later section.

In the second step of automatic gain equalization, networks are provided to match system gain variations caused by electron tube aging and by changes in repeater hut temperatures. The loss characteristics of these equalizers are controlled by thermistors, which in turn are controlled by the 308-kc and the 2,064-kc pilots. These equalizers are used every 40 to 120 miles.

In the final step of automatic gain equalization, networks are provided to compensate for second-order effects of the first three rapid variations described above, namely, cable loss variations, and repeater gain variations caused by hut temperature changes and electron tube aging. The loss characteristics of these networks are under control of thermistors acted on by the 556-kc, 3,096-kc, and 8,320-kc pilots. These equalizers are located at approximately 150-mile intervals.

The thermistors which control the loss-frequency characteristics of automatic equalizers are driven by regulators through a simple form of analogue computer. The design and operation of this circuit is described in a companion paper.[8]

There is no automatic control of the delay characteristic in the system except that provided by the automatic gain equalizers. Every effort is made to have these equalizers match the transmission changes outside the band so that resulting delay changes in the band are minimized.

Equalization System Considerations

Whether the system is being equalized for telephone or television, it is im-

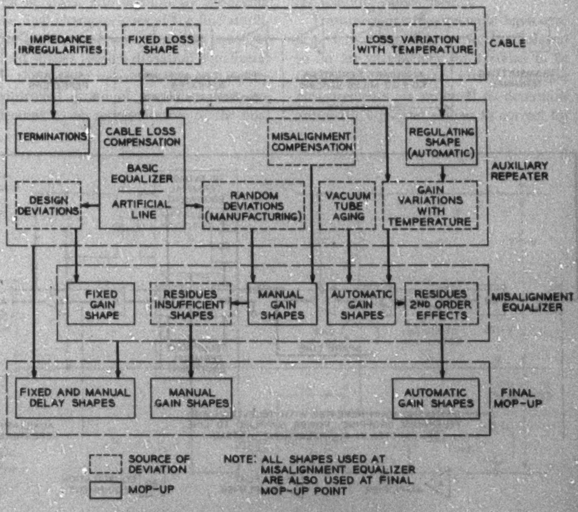

IMPEDANCE IRREGULARITIES
FIXED LOSS SHAPE
LOSS VARIATION WITH TEMPERATURE
CABLE
TERMINATIONS
CABLE LOSS COMPENSATION
BASIC EQUALIZER
ARTIFICIAL LINE
MISALIGNMENT COMPENSATION
REGULATING SHAPE (AUTOMATIC)
AUXILIARY REPEATER
DESIGN DEVIATIONS
RANDOM DEVIATIONS (MANUFACTURING)
VACUUM TUBE AGING
GAIN VARIATIONS WITH TEMPERATURE
FIXED GAIN SHAPE
RESIDUES INSUFFICIENT SHAPES
MANUAL GAIN SHAPES
AUTOMATIC GAIN SHAPES
RESIDUES 2ND ORDER EFFECTS
MISALIGNMENT EQUALIZER
FIXED AND MANUAL DELAY SHAPES
MANUAL GAIN SHAPES
AUTOMATIC GAIN SHAPES
FINAL MOP-UP
SOURCE OF DEVIATION
MOP-UP
NOTE: ALL SHAPES USED AT MISALIGNMENT EQUALIZER ARE ALSO USED AT FINAL MOP-UP POINT

Fig. 13. L3 coaxial system-equalization plan

mediately apparent that the channel requirements described earlier, which were applied after 4,000 miles of transmission imply that, with no equalization, stability of the transmission characteristics of the individual repeaters would have to be of the order of a few ten thousandths of a decibel. Obviously, stabilities of this magnitude with changes caused by temperature, electron tube aging, and manufacturing processes cannot be achieved. Therefore, the equalization system design must be based on an economical balance between the cost of achieving repeater accuracy and stability and the cost of providing and maintaining an elaborate system of fixed, manual, and automatic equalizers.

The equalization problem involves so many variables that no attempt has been made to evolve a unified theoretical basis for evaluating the factors entering into this economic balance. However, in planning and designing the *L3* system a number of principles and points of view have been developed which have guided the equalization planning.

Misalignment. The transmission objectives described above are determined on the basis of delivering satisfactorily equalized signals at terminals. In addition to this function, the equalizers must limit the signal excursions along the line so that excessive noise or modulation is not accumulated in the repeater system. The amount of signal misalignment that can be allowed to accumulate before the first mop-up equalizer depends of course on the signal-to-noise allowance that has been made for this purpose. The amount of signal-to-noise performance allotted to misalignment must represent a balance between the reduced repeater spacing and increased complexity of equalizers that it costs and the increased spacing between mop-up equalizers and increased repeater deviations that it allows.

The engineering method for arriving at this balance represents an interesting example of system design by successive approximations. For example, the total gain area available over an infinite frequency range, in a coupling network is inversely proportional to the capacity across the network and one of the important design choices is the extent to which one tries to utilize this area in the transmitted frequency band. The degree to which the available gain is concentrated in-band is called the resistance integral efficiency. In the very early stages of the amplifier design it was necessary to choose resistance integral efficiencies and frequency characteristics for the coupling networks. In a definite but complicated way these parameters are related to the sensitivity of the networks to element variations. Efficient networks give improved signal-to-noise performance but also increase the sensitivity to element changes. By examining deviation curves for a number of specific network designs, tentative choices were made of 50-per-cent resistance integral efficiency for the coupling networks and an allocation of about half the cable slope to the pair of coupling networks and the remainder to the feedback network. With an amplifier employing these networks a detailed study was made of the noise and modulation penalties at two frequencies resulting from misalignment in several lengths of line. This study indicated that with certain refinements in the repeater design the misalignment in 20 or more auxiliary repeater sections could be tolerated with a signal-to-noise penalty of about 2 decibels which was judged to be a reasonable allotment for this purpose. In addition, this study brought out: 1. that randomizing the variations of an element between its normal manufacturing limits resulted in a 4-to-1 reduction in the required misalignment allowance as compared with accepting large numbers of elements at one extreme of their limit; and 2. that a small amount of gain adjustment at each repeater in the vicinity of the high magnitude television carrier would reduce the required misalignment allowance by about 2-to-1. Refinements on this plan for handling misalignment had to wait until the signal-to-noise and repeater design were crystallized. However, this study provided a powerful tool for evaluating proposed element deviations during the design period.

When the exact signal levels and the most limiting modulation products became known and when the repeater

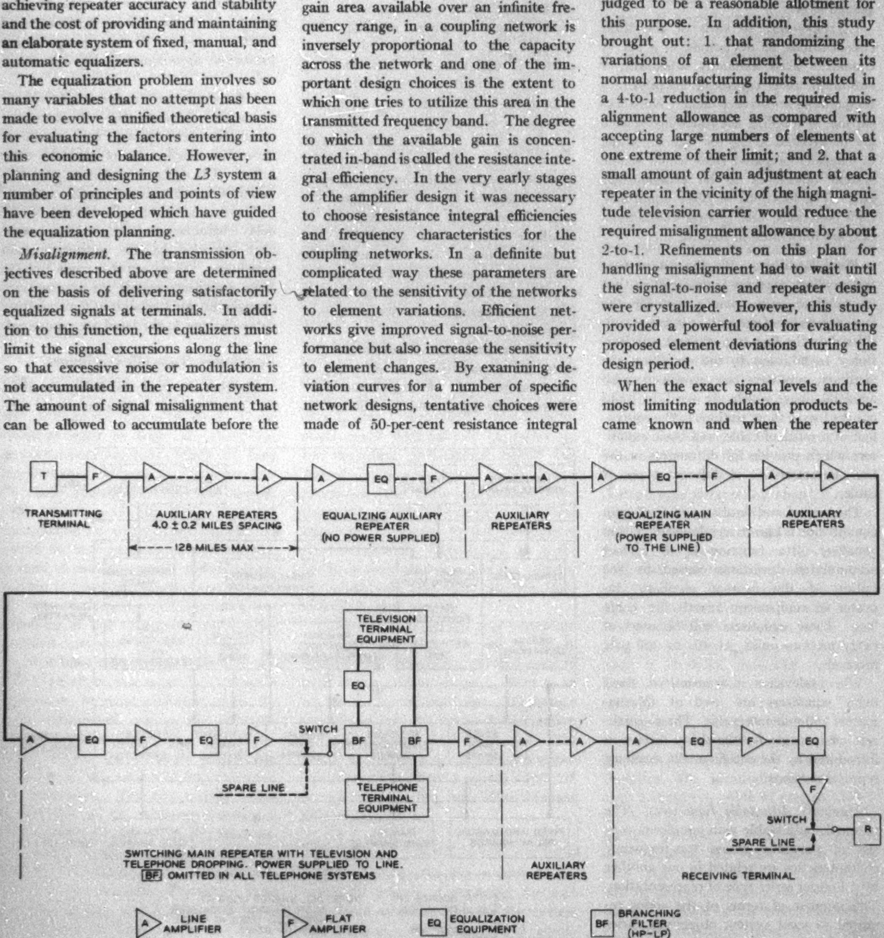

T
F
A
EQ
TRANSMITTING TERMINAL
AUXILIARY REPEATERS 4.0 ± 0.2 MILES SPACING
128 MILES MAX
EQUALIZING AUXILIARY REPEATER (NO POWER SUPPLIED)
AUXILIARY REPEATERS
EQUALIZING MAIN REPEATER (POWER SUPPLIED TO THE LINE)
AUXILIARY REPEATERS
TELEVISION TERMINAL EQUIPMENT
SWITCH
BF
SPARE LINE
TELEPHONE TERMINAL EQUIPMENT
SWITCHING MAIN REPEATER WITH TELEVISION AND TELEPHONE DROPPING. POWER SUPPLIED TO LINE.
BF OMITTED IN ALL TELEPHONE SYSTEMS
AUXILIARY REPEATERS
SWITCH
SPARE LINE
R
RECEIVING TERMINAL
A LINE AMPLIFIER
F FLAT AMPLIFIER
EQ EQUALIZATION EQUIPMENT
BF BRANCHING FILTER (HP-LP)

Fig. 14. General layout of L3 repeaters

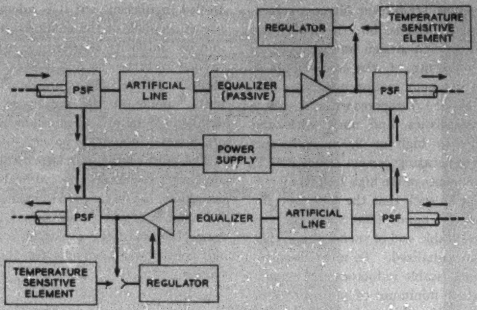


Fig. 15. Auxiliary repeater

characteristics and final element deviations were determined, it became possible to make a refined study of misalignment in terms of the noise and modulation impairment associated with specific signals and distortion products. At this point, performance margins associated with specific interferences could be used to allow more or less misalignment of the particular signal components forming the interference. Likewise, amplifier deviations with specific frequency characteristics could be evaluated exactly in terms of their effect on the number of repeaters between mop-up equalizers. By studies of this type it was determined that the *A* or misalignment equalizers could be spaced at intervals not to exceed 32 auxiliary repeater sections.

Distribution of Element Deviations. The methods of statistical quality control used to monitor the process of manufacture provided the necessary techniques for obtaining the desired randomization of deviations. In the techniques that were developed to apply the broad field of knowledge on quality control to the specific needs of the *L3* system, the most important point to be appreciated is that the control of the process of manufacture (as well as the end electrical requirements) of individual elements is being used as a basic factor in the design of the system.[12]

Repeater Accuracy. In developing the equalization plan, it is a logical and straight-forward operation to provide shapes and ranges in the equalizers that will compensate for the random variations of known elements. Likewise, real but indeterminate parasitic elements can be taken into account by specifying the final characteristics of the line amplifier feedback network and the equalizer fixed shapes (design deviation equalizers) on the basis of measurements on a rigidly controlled group of amplifiers that are deemed to be representative of the final product. However, having once specified the equalization on this basis, the design elements and indeterminate parasitic elements must be held to the values and ranges upon which equalizer location, shapes, and ranges are specified. This point of view has led to rigid mechanical control and the omission of component adjustments in the line amplifier which represent a departure from other transmission systems. These features are discussed in detail in a companion paper on amplifiers.[9]

NEW YORK-PHILADELPHIA TRIAL

The first installation of *L3* has been made between New York and Philadelphia. Since the middle of 1952, this installation has been used to test components, to verify values of important system parameters used in system analyses, and to gather data for the further design and development of equalizers.

Random noise measurements have confirmed theoretical values, Fig. 5, to an accuracy of better than 2 decibels. In general, the measurements have indicated that the theoretical values have been conservative.

Measurements of system modulation performance, made with single frequency tones, also confirm the theoretical values used in the analyses. Third-order modulation measurements are in almost complete agreement with theory while second-order measurements have been generally two to three decibels more favorable than the analytic values used.

Transmission measurements have confirmed that equalizer networks designed so far are satisfactory for systems to be installed in the near future. Further measurements are required to determine automatic equalizer shapes to correct for

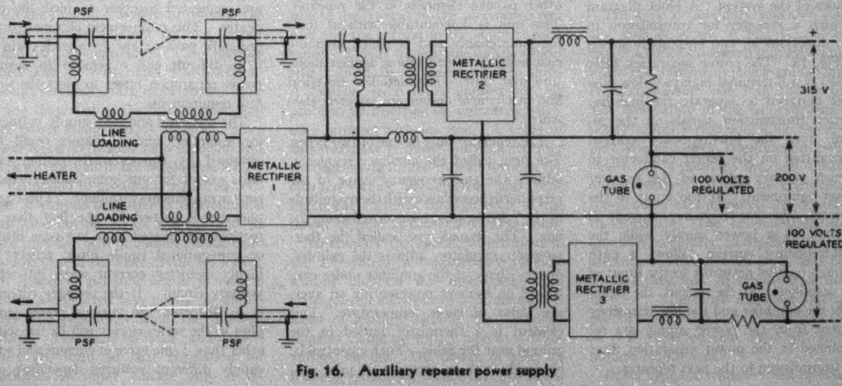


Fig. 16. Auxiliary repeater power supply

the second-order effects of temperature changes and electron tube aging. Also under active study are the problems associated with final mop-up for long television systems.

System Description

General

In the preceding sections on system design the functions of the auxiliary repeaters and the need for additional repeaters with varying amounts of equalization have been brought out. Fig. 14 shows the transmission layout of a typical *L3* system. The auxiliary repeaters contain amplifiers and regulating equipment to compensate for the basic cable loss and its variation with temperature. Since such repeaters are dependent on the cable for their primary source of power, they are called auxiliary repeaters.

At points in the system where additional first-order equalization is required to reduce misalignment the complexity of the repeater equipment increases and such repeaters receiving power over the cable are called equalizing auxiliary repeaters.

The distance which power may be transmitted over the cable to the auxiliary repeaters is limited; therefore, repeaters at specified intervals must be capable of supplying power to the cable. These are called main repeaters. They may be equalizing main repeaters where only first-order equalization is required or switching main repeaters where lines are switched or circuits dropped.

Auxiliary Repeater

Transmission Circuit

This repeater is the basic unit of the system and its design determines to a great extent the performance and economics of the system. A block diagram of such a repeater for transmission in two directions on two coaxials is shown in Fig. 15. The power separation filter (PSF) is a 6-terminal high-pass–low-pass filter designed to separate the high-frequency transmission signals on the coaxial from the low-frequency current transmitted on the center conductor to furnish primary power to the repeater power equipment. At the input to the repeater, the low-frequency current is diverted to a power supply while the high-frequency current follows a path through passive networks to the input of the amplifier. At the output, the signal from the amplifier and the low frequency current from the power supply are recombined in the power separation filter for transmission to the next repeater.

The power separation filters are basically simple designs, but the realization of the theoretical design was complicated by the following: 1. the components in the low-frequency section must pass currents of about 1.5 amperes without change in characteristics, and must withstand potentials as high as 2,000 volts rms without generating corona noise, and 2. the components in the high-frequency section must be such that the loss over the transmission band (300–8,350 kc) is small and stable and of such a shape that it is easily equalized. To meet these requirements, stable inductors and capacitors with a minimum of parasitic resonances in the band were designed.

The artificial line shown preceding the input to the amplifier is a passive network to build out the loss-frequency characteristic of a short cable section to be equivalent to the loss-frequency characteristic of 4.0±0.2 miles of 0.375-inch cable or 2.87±0.15 miles of 0.27-inch cable. These lines are provided in several different sizes so that, where it is impossible physically to locate the repeater within the specified accuracy of 0.4 mile, this accuracy can be obtained electrically. The design of the network is such that an accurate and stable characteristic is obtained with a minimum number of components.

The equalizer is a means for compensating for small variations in the transmission characteristics of coaxial cables which are caused by variations in the physical construction of the cable. In the case of the most generally used cable, this equalizer inserts only a small flat loss.

The amplifier is of the feedback type whose gain-frequency characteristic is closely equivalent to 4.0 miles of 0.375-inch coaxial cable plus the loss of the other passive elements in the repeater. This unit is demountable without tools for maintenance and is sealed in a die cast housing as protection against moisture and dust. The detailed electrical and mechanical design are covered elsewhere.[9]

The regulator may be one of two types. The first, called the auxiliary regulator, adjusts the gain-frequency shape of the amplifier in accordance with the magnitude of the 7,266-kc pilot transmitted along the line. The second type, called the thermometer regulator, adjusts the gain-frequency shape of the amplifier under control of an element representing an average value of cable temperature. This element is a thermistor buried in the ground near the cable. Such a control is, obviously, not as accurate as pilot-controlled regulation, but it is adequate for use at one-half of the auxiliary repeaters and its simplicity results in considerable saving in first cost and power requirements. The regulators are demountable units similar to the amplifiers. Their detailed electrical and mechanical design are covered in another paper.[8]

The pilot alarm unit is provided with auxiliary regulators to indicate pilot deviations beyond a predetermined limit. Its operation will be described later in connection with the discussion of alarm and control arrangements for the entire system.

Power Supply

Primary a-c power for the auxiliary repeater is supplied on a constant current basis from the main repeater over the center conductors of the two associated coaxials. Power generating and control equipment used at the main repeater will be discussed under "Power Generation and Transmission." At the auxiliary repeater, power supply equipment is required to convert the primary power to suitable voltages for heater, plate, and bias use as shown in Fig. 16. Half of the input to the power supply is taken from each center conductor and the output of the power supply is used to power the entire 2-way repeater.

The heater voltages are obtained by simple transformation which is complicated only by the fact that accurate and low-loss transformers are required and that the primaries of these transformers must withstand high a-c potentials without generating corona noise which might be transmitted through the power separation filter to the input of the amplifier. Two separate transformers are used to split the load between the two center conductors, even though the secondaries are connected together to feed the repeater. This arrangement eliminates one crosstalk path at low frequencies where it is difficult and expensive to design power separation filters to meet the system requirements.

The d-c plate and bias supply voltages for amplifiers and regulators could be obtained by conventional rectifier circuits except for one complication which such arrangements introduce. This complication is caused by the fact that a rectifier terminated in a low-pass filter, or conventional ripple filter, reflects a highly distorted current wave into the primary circuit. If the primary current is so distorted the various power supplies in the series circuit will be fed with other than a sine wave of current and will supply different voltages depending on

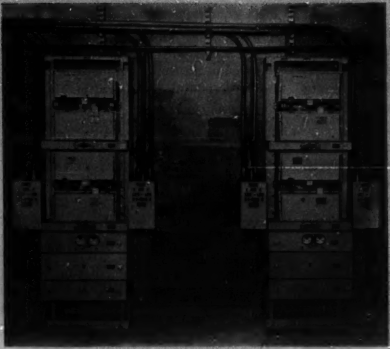


Fig. 17 (above). Typical auxiliary repeater in concrete block hut

Fig. 18 (right). Typical main repeater installation

the wave form. Since the heater power depends on the rms value of current while the d-c output depends on the peak value of voltage, it is easily seen that the relationship between these two will change with the wave form of the applied current. Furthermore, the line loading to be discussed later must be calculated on the basis of a pure sine wave; appreciable harmonics in the line current tend to make it impossible to predetermine the loading to any reasonable degree of accuracy.

It was found that these problems could be avoided and the power factor of the power supply made very nearly unity if the rectifier (RECT 1) was terminated in a constant resistance load rather than a low-pass filter. This was provided by paralleling the conventional low-pass filter with a high-pass section terminated in the proper resistance load. To avoid wasting the power in this load, a second rectifier was added (RECT 2). The d-c output of this circuit is used in series with the main d-c supply to provide the higher voltage required for the output stage of the amplifier. This rectifier must also be terminated resistively although its effect on the main current wave is less than that of the first rectifier, and the power dissipated is smaller. Since there was a further use for a small amount of power for bias in the regulators, RECT 3 was added to produce a regulated voltage in conjunction with a conventional gas tube circuit. This rectifier and its load provide the termination for the high-pass section of the filter circuit for RECT 2. A second gas tube circuit is used to obtain a regulated bias supply for amplifiers and regulators from the 315-volt source. The loads on both gas tube circuits are fixed so that regulation for variation in input voltage only is required. For this reason a low-current highly stable gas tube could be used.

Power Loading

The power transmission circuit of a power loop is essentially a resistance-capacity network at the power frequency. The line and the power supply resistance are the resistance component, and the line and power separation filter capacity to ground make up the shunt capacity. If the circuit were used in this form, the primary current at each repeater would be different since it would be the vector sum of the current in the succeeding section and the current in the shunt capacitance. This is undesirable as the objective is to make all power supplies alike. A familiar solution is applied to this problem by inserting an inductive reactance in series with the line. A value of this reactance is chosen for each repeater to compensate for the current through the effective shunt capacitance, and thus make the currents through each of the power supplies as nearly alike as possible.

To simplify the loading adjustment in the *L3* system, a continuously variable loading inductor was developed. This arrangement allows more accurate adjustment of loading without the complications of changing wiring taps in a high-voltage circuit. The design of such an inductor presented formidable obstacles, as a large range of variation was desired (20 to 120 millihenrys) and relatively high currents and voltages involved. The device used consists of two inductors which may be rotated with respect to each other, so that the coupling between their magnetic circuits varies ideally between zero and 100 per cent. One inductor is inserted in each side of the power circuit and a net result is obtained which is equivalent to varying each inductor.

Physical Description

The type of auxiliary repeater generally used is shown in Fig. 17. It consists of a 6-foot cable duct framework upon which the component panels are mounted. It is completely wired in the factory. The lower third of the bay contains the power supply equipment while the upper part contains two transmission panels. One panel is provided for each direction of transmission and all of the transmission components of the circuit are found on these panels. The demountable units, amplifier, regulator, and pilot alarm unit are interconnected with plugs and jacks so that they may be removed for maintenance. The other units are interconnected by screw-type terminals and cable as it is expected that they will seldom require maintenance.

Other types of repeaters will be available to meet special conditions such as manholes where sealed apparatus cases will be required to prevent damage due to water submersion, or telephone offices where standard 11-foot 6-inch frameworks are usually desired.

EQUALIZING AND SWITCHING REPEATERS

Components

The equalizing auxiliary and main repeaters use the same general types of transmission equipment as the auxiliary repeater except for the equalizers. They differ principally in the quantity of equalization equipment provided, and the bay arrangements. Table III lists a summary of the basic transmission units in each repeater.

At these repeaters a line amplifier is used as a receiving amplifier to compensate for the previous section of cable. Flat amplifiers are used as transmitting amplifiers and to compensate for the loss introduced by the equalizers. They have a flat gain-frequency characteristic and no provision for pilot control of their gain. Their design is very similar to that of the line amplifier.[9]

The functions of the fixed, the manual, and the pilot-controlled equalizers are noted in the section on "The Mop-up Plan" of this paper and are discussed in detail elsewhere.[8]

Equalizing Auxiliary Repeaters

This type of repeater will be found after a maximum of 32 auxiliary repeaters provided power feed to the cable, dropping, or switching is not required, see Fig. 14. The major components provided are covered in Table III. In addition to these items, power separation filters, basic equalizers and artificial lines identical with those in auxiliary repeaters are used. A pilot alarm unit is also included to monitor each of the three regulators and transmit an alarm when any one of the controlling pilots has deviated beyond a given limit.

Power for these repeaters is obtained from the cable just as in the case of the auxiliary repeater, and much of the same type of equipment is used. However, because of the larger amount of power required, and the layout of the repeater, the auxiliary repeater power units have been repackaged to provide the optimum arrangements for leads carrying high current or critical bias supply circuits.

The design of panels used in this repeater was dictated by the general scheme conceived for the switching main repeater where the maximum amount of equipment is required. This arrangement involves the use of both sides of a duct-type frame. A single panel, again called the transmission panel, is used, but an amplifier is mounted on one side and a regulator on the other. This requires access to both sides of the bay, but results in an over-all saving in the number of bays and over-all floor space.

All of the transmission components and a heater and bias supply unit are mounted in one 7-foot bay for each coaxial. The plate and primary a-c power for two of these bays is mounted in another 7-foot bay.

Equalizing Main Repeater

This repeater contains exactly the same transmission equipment as the equalizing auxiliary repeater; see Table III. It differs in the function noted before, that is, it is equipped to feed power to the cable. The equipment to perform this function will be described later. Since the repeater can feed power to the cable, it can also supply the power for its own operation. This power is derived from the primary a-c supply used for the line by means of conventional metallic rectifiers for d-c circuits and transformers for the a-c heater supplies. These power supplies are not a part of the power loop containing auxiliary repeaters, so no special arrangements are required to obtain good waveform or high power factor.

The equipment arrangement uses the same units as the equalizing auxiliary repeater, but here conventional 11-foot 6-inch frames are used.

Switching Main Repeater

Usually this type of repeater is supplied at the point where circuits are dropped or terminated. To permit switching from a working line to a spare line in case of trouble, (see the section entitled "Automatic Switching") more complex equalization is required so that the lines will be as nearly alike as possible when the switch takes place. Furthermore, the signal delivered to the terminal must meet equalization limits that will result in a satisfactory grade of service.

Where the repeater is part of a system required to transmit only message circuits the basic equipment shown in Table III (telephone only case) is required. In addition to these units, facilities are provided which indicate and alarm pilot levels and provide for patching and other maintenance arrangements. Since this repeater always feeds power to the cable, it uses the same power arrangements as the equalizing main repeater.

Where the system is being used for the combined telephone and television signal this repeater is the same as the all-telephone system repeater except that it has additional equalization equipment to adjust the system for the more stringent television requirements. Furthermore, line-connecting equipment consisting of branching filters and additional equalization is required. Branching filters are used to separate the telephone and television bands so that they can be transmitted to their respective terminal equipment. These are combined high-pass—low-pass structures complicated by strict requirements on stability of the gain-frequency and delay-frequency characteristics. Delay equalization for the line sections must also be provided in the television branch and a large part of this is combined with the branching filters. Other components required for long television systems are adjustable gain and delay equalizers and associated amplification.

The same type of equipment is used as that described for the other repeaters except that a number of additional transmission panels are required to mount the additional amplifiers and regulators associated with the equalizers. Two 11-foot 6-inch bays are used to contain the equipment for one through coaxial. One bay contains the receiving equipment which precedes the line switch. The other bay contains the transmitting equipment

Table III. Summary of High-Frequency Line Equipment at Repeater Stations

	Auxiliary	Equalizing Auxiliary	Auxiliary Main	Switching Main (Telephone Only)	Switching Main Telephone-Television
Line amplifier	1	1	1	1	1
Flat amplifier	0	2	2	5	5
Auxiliary or thermometer regulators	1				
Manual equalization (no. of terms)		10	10	15	25
Automatic equalizers	1	3	3	6	6
Regulators for equalization, kc		7,266 308 2,064	7,266 308 2,064	7,266 308 2,064 3,096 556 8,320	7,266 308 2,064 3,096 556 8,320
Design deviation equalizer		1	1	1	1

(transmitting amplifier and hybrid) and any line connecting equipment for combined systems. Fig. 18 shows a typical main repeater installation.

AUTOMATIC SWITCHING

To preserve transmission in the event of the failure of a component of the system and for transmission maintenance purposes, one coaxial in each direction is operated as a stand-by. An automatic switching system is provided to permit substitution of the stand-by line for any of the working lines. The lines are switched at the input to the transmitting amplifier and at the output of the receiving amplifiers and equalizers as shown in Fig. 14.

At the receiving end of a switching section, equipment is provided whose function is to recognize failure of a working line and initiate the switching circuits. Information as to the transmission conditions of the system exists in the pilot regulators, the output of which controls a sensitive relay with high- and low-limit contacts. The operation of one of these relays provides the switching system with the information that transmission has failed or been seriously impaired. It is necessary to make a switch as rapidly as possible in the case of a total failure in order to reduce the effect upon the transmission circuits. As the relays take appreciable time to operate, the receiving switch equipment is designed to operate directly from the d-c output of the regulator associated with the 7,266-kc line pilot. This permits complete switch operation in about 15 milliseconds.

Upon receipt of information from the regulators that one or more of the pilots have gone out of limits, the switch initiator signals the transmitting switch control equipment at the transmitting end of the switching section as to which line has failed. Signalling is accomplished by the use of tones in the 280- to 296-kc range which are transmitted over all coaxials in parallel in the reverse direction. Use of the coaxials for transmission of these signals obtains maximum speed of operation of the switch. All paths are used in parallel to preclude failure of the switch if one channel in the opposite direction would be inoperative.

The transmitting switch control equipment causes the transmitting end of the stand-by line to be switched in parallel with the line in trouble and then signals the switch initiator that this has been done. This verifier tone actuates the line switch at the receiving end to complete the switch.

When the trouble clears, the switch is released as the initiator checks every minute to see if service on the working line can be restored. In the event of prolonged trouble, the switch can be locked manually and the initiator will no longer attempt to restore to normal. Release of the switch is accomplished by the transmission of a release tone to the transmitting end while a checking tone returned by the transmitting end indicates completion of release and readies the switch initiator for further switching.

For maintenance purposes, manual operation of the switching equipment is provided. In effect, manual switches are made by simulating a failure. Alarm features are provided to indicate to the operating personnel failure of the coaxial system or failure of the switching equipment. Care has been taken in the design to insure that failure of the switching equipment in no way affects transmission except by removing the protection afforded by the presence of the switching facility.

One 11-foot 6-inch bay is required for the switch control equipment for each direction of an 8-coaxial system. The

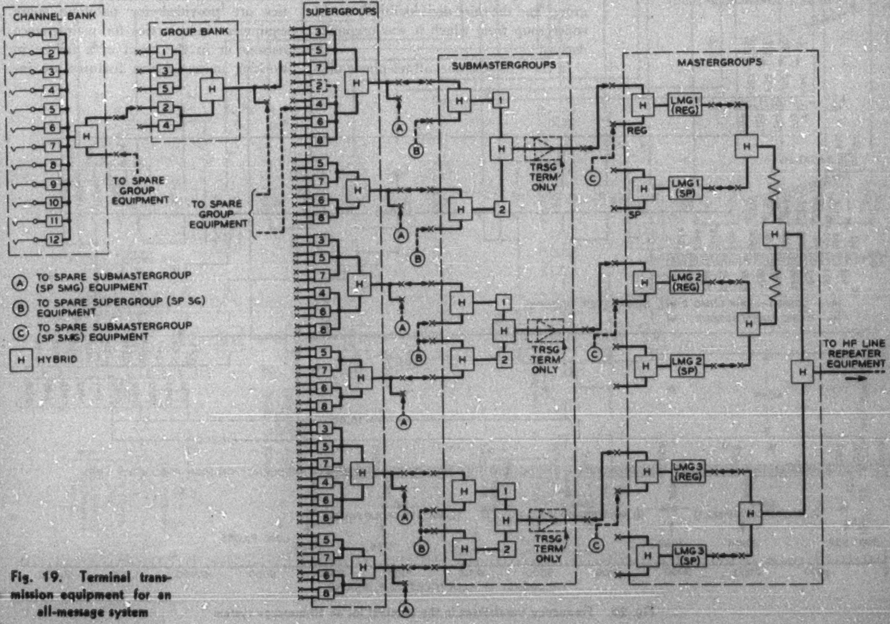


Fig. 19. Terminal transmission equipment for an all-message system

line switches are mounted in the miscellaneous bay of the main repeater lineup.

TERMINALS

Television terminal equipment, which is required to modulate the video frequency signals to and from the high-frequency line, is described in a companion paper[10] and therefore will not be discussed here.

The telephone terminals consist of modulators (and related transmission equipment) and carrier and pilot generating equipment. The transmission components of a terminal for an all message system are shown in Fig. 19. The channel, group, and supergroup equipment are designs previously used in the *L1* system. The designs of the submastergroup and mastergroup units employ circuit arrangements similar to those used in the super group equipment. The greater bandwidths, higher frequencies, and more severe stability requirements demanded new components and improved circuit and layout techniques.

Fig. 20 shows the modulation steps and location in the frequency spectrum of the supergroups, submastergroups, and mastergroups when the *L3* system is used for telephone and television or all telephone. Mastergroup 1 comprises the first ten 60-channel supergroups. This mastergroup is placed directly on the line in the 564 to 3,084-kc frequency band for both the telephone-television and all-telephone cases. When the system is used entirely for telephone, two additional mastergroups are formed by modulating mastergroup 1 up into the desired frequency bands.

Mastergroup 1 is subdivided into two submastergroups. The lower six supergroups, comprising submastergroup 1, are modulated directly up from the basic supergroup located in the 312 to 552-kc band. The modulation and carrier supply equipment for these supergroups are the same units that are employed in the *L1* system. The upper four supergroups comprising submastergroup 2 are obtained by modulating four supergroups located in the same frequency range as the top four supergroups in submastergroup 1 into the top part of mastergroup 1.

The supergroup numbering system used for *L3* has been adopted for easy identification of supergroups in their high-frequency positions. Each supergroup is given a 3-digit number. The first digit identifies the mastergroup, the second digit identifies the submastergroup, and the third digit identifies the *L1* supergroup from which it was originally derived.

In the 1,860-channel all-telephone allocation, supergroup 112, which corresponds to the basic *L1* supergroup 2, may be used for high-quality, long-haul message circuits. When the system is used for telephone and television, supergroup 112 is restricted to circuits under 200 miles in length because of intolerable second-order cross-modulation between these signals and the television signal.

With these groupings of channels, new modulating and carrier supply equipment is required for submastergroup 2 and mastergroups 2 and 3. The frequency allocation and modulation steps shown in Fig. 20 were chosen so that good performance could be obtained with relatively inexpensive filters used for suppressing unwanted side bands and to develop lower side-band signals appearing right side up for transmission over the high-frequency line. The carrier frequencies used in the modulation steps were chosen so that some of the same filters could be used in submastergroup 2 and mastergroup 2 and 3 modulators. The clear bands between the submastergroups and the mastergroups were chosen to permit the use of economical blocking filters at repeaters where circuits are to be dropped from the high-frequency line and to provide frequency space for the line pilots.

It will be noticed in Fig. 19 that facilities are provided for patching spare equipment into service for maintenance reasons or in the event of a failure in working units. Alarm features are in-

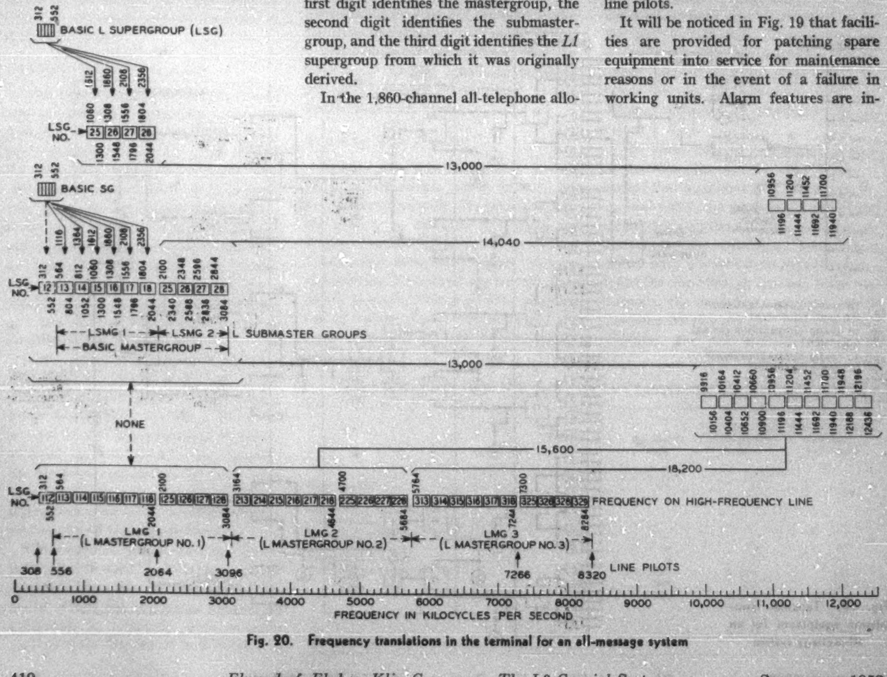

312
552
BASIC L SUPERGROUP (LSG)
1612
1860
2108
2356
1060
1308
1556
1804
LSG NO.
25
26
27
28
1300
1548
1796
2044
13,000
10956
11204
11452
11700
11196
11444
11692
11940
BASIC SG
1116
1364
1612
1860
2108
2356
14,040
312
564
812
1060
1308
1556
1804
2100
2348
2596
2844
12
13
14
15
16
17
18
552
804
1052
1300
1548
1796
2044
2340
2588
2836
3084
LSMG 1
LSMG 2
L SUBMASTER GROUPS
BASIC MASTERGROUP
13,000
9916
10164
10412
10660
10956
11204
11452
11700
11948
12196
10156
10404
10652
10900
11196
11444
11692
11940
12188
12436
NONE
15,600
18,200
312
564
2100
3164
4700
5764
7300
112
113
114
115
116
117
118
125
126
127
128
213
214
215
216
217
218
225
226
227
228
313
314
315
316
317
318
325
326
328
329
FREQUENCY ON HIGH-FREQUENCY LINE
552
2044
3084
4644
5684
7244
8284
LMG 1 (L MASTERGROUP NO. 1)
LMG 2 (L MASTERGROUP NO. 2)
LMG 3 (L MASTERGROUP NO. 3)
308
556
2064
3096
7266
8320
LINE PILOTS
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10,000
11,000
12,000
FREQUENCY IN KILOCYCLES PER SECOND

Fig. 20. Frequency translations in the terminal for an all-message system

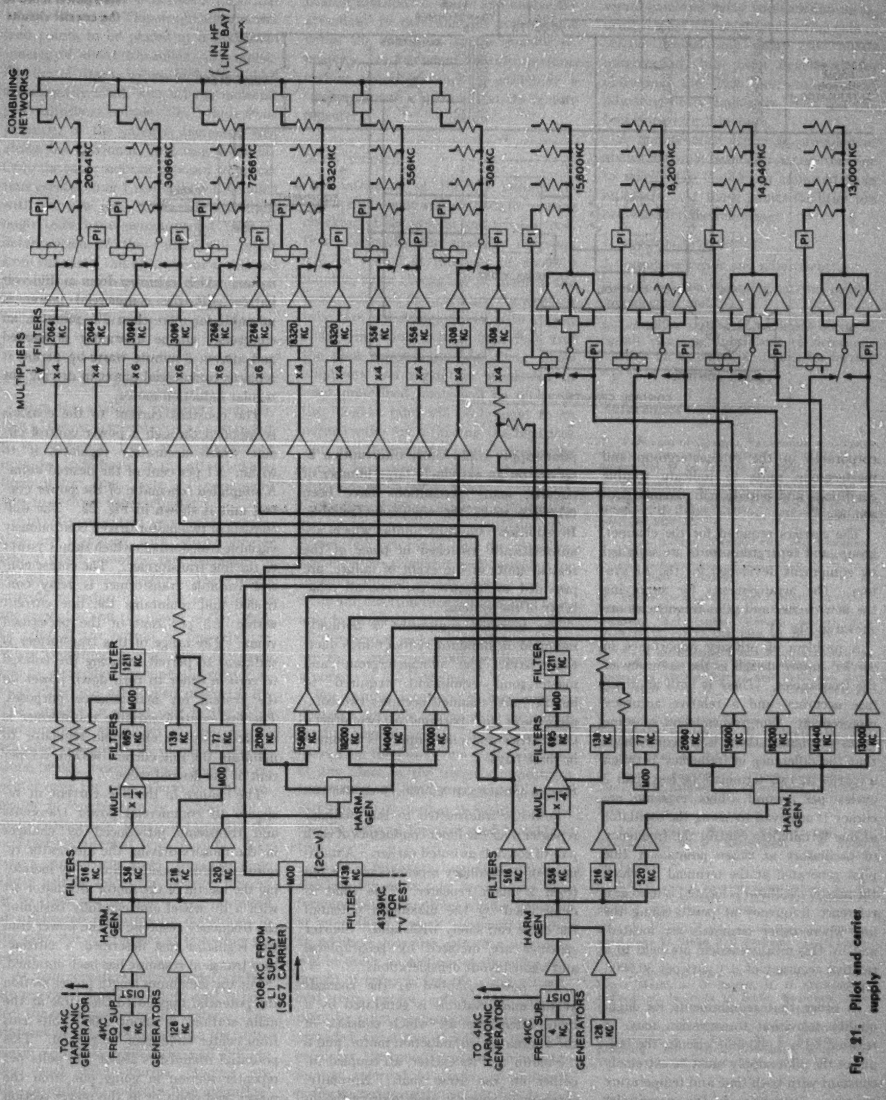

COMBINING NETWORKS
IN HF LINE BAY
X
2064KC
3096KC
7266KC
8320KC
556KC
308KC
15,600KC
18,200KC
14,040KC
13,000KC
PI
MULTIPLIERS
FILTERS
2064 KC
3096 KC
7266 KC
8320 KC
556 KC
308 KC
x4
x6
FILTER
1211 KC
MOD
FILTERS
695 KC
139 KC
77 KC
MULT
x 1/4
2080 KC
15600 KC
18200 KC
14040 KC
13000 KC
HARM. GEN
516 KC
556 KC
216 KC
520 KC
(2C-V)
4139 KC
FILTER
4139KC FOR TV TEST
2108KC FROM L1 SUPPLY (SG7 CARRIER)
TO 4KC HARMONIC GENERATOR
4KC FREQ SUP
DIST
4 KC
128 KC
GENERATORS

Fig. 21. Pilot and carrier supply

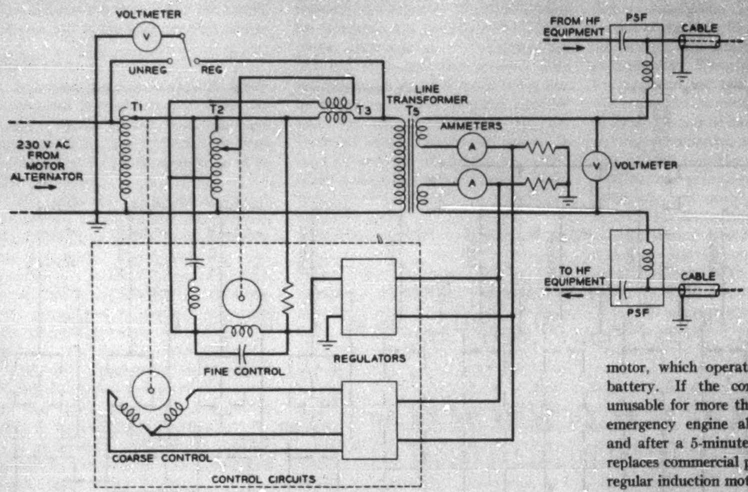


Fig. 22. Simplified schematic of the power control circuit through which regulated power is fed to the coaxial circuits

corporated in the submastergroup and mastergroup units to indicate trouble conditions and initiate maintenance procedures.

The carriers required for the channel, group, and supergroup units are supplied by equipment developed for the *L1* system. The arrangements for supplying the new carrier and pilot frequencies are shown in Fig. 21.

A problem of primary importance in carrier supply design is the accuracy of the frequencies. There is both an absolute accuracy and a relative accuracy requirement. For transmission of some types of signals there is a requirement that the difference in frequency between a carrier at two terminals be less than 2 cycles per second. This extreme accuracy is achieved by using the oscillator at one terminal to control the frequency of oscillators at other points. A line pilot generated at the terminal in which the master oscillator is located is used as a reference frequency at points along the line where other terminals are located, and by this means carriers are held to a relative accuracy of ±1 part per 30,000,000.

In order that requirements for high-quality television transmission may be realized on a 4,000-mile circuit, the output of the pilot supply must be extremely constant with both time and temperature changes. Deviations in the magnitudes of line pilots are maintained to less than 0.05 decibel.

Since a failure in the *L3* carrier and pilot supply could cause interruption to service on an extremely large number of circuits, many precautions have been taken to make the equipment reliable. In addition, stand-by units, which are automatically switched in place of the regular units in the event of failure, are provided to improve the over-all reliability of the system.

The terminal equipment is normally mounted on standard 11-foot 6-inch duct-type bays. The submastergroup and mastergroup equipment required to handle 1,860 channels occupies two complete bays and portions of two others. One carrier and pilot supply is mounted in three bays.

POWER GENERATION AND TRANSMISSION

Power is transmitted to the auxiliary repeater over the inner conductors of each pair of coaxials as noted earlier. A maximum of 21 auxiliary repeaters can be fed from a main repeater. This limit is determined by the maximum potential the cable can safely withstand. Shorter spacings are dictated by geographical and plant layout considerations.

The power supplied to the coaxials at the main stations is generated by a motor-alternator set which consists of the alternator, an induction motor, and a d-c motor and its exciter, all coupled together on the same shaft. Normally, commercial power is used to drive the induction motor. When this source fails or the voltage goes out of prescribed limits, the drive is transferred to the d-c motor, which operates from a 130-volt battery. If the commercial power is unusable for more than 2$^{1}/_{2}$ minutes, an emergency engine alternator is started and after a 5-minute warm-up period it replaces commercial power in driving the regular induction motor.

The constant current to the coaxials is supplied through a power control circuit which accurately regulates it to within ±1 per cent of the desired value. A simplified schematic of the power control unit is shown in Fig. 22. The unit consists of two motor-driven continuously variable transformers which supply power to the line transformer. The course control variable transformer is relay controlled and maintains the line current within ±3 per cent of the prescribed value. The range of this transformer is sufficient to permit reducing the voltage to zero in order to turn down power on the system for maintenance purposes. The fine control variable transformer is regulated by an electronic regulator to maintain the line current within ±1 per cent of the desired value.

The change in the line current in response to commercial power transients and transients introduced by changes in the motors driving the alternator requires careful consideration. By increasing the inertia of the motor alternator set with a fly wheel and carefully designing the frequency response of the power control regulator just described, a satisfactory transient response has been obtained.

For the maximum length power section the potential applied to the cable at the main stations is about 2,000 volts rms from center conductor to ground. This potential diminishes about 100 volts per repeater section in going out from the power feed point or as the power section is shortened. (The maximum potential applied to the cable in *L1* systems is about 800 volts rms between center conductor and ground). Extensive tests on the installed cable showed that corona develops in the cable at potentials varying in a random fashion between 1,200 and 1,600 volts rms. This would allow power feed points to be placed at a maximum spacing of about 100 miles. By replacing the nitrogen, with which the cables are normally filled, with a large molecule gas, sulfurhexafluoride (SF_6), the corona potential of the cable is increased well above the maximum operating potential. Only the cable sections exposed to potentials greater than 1,200 volts will be filled with the new gas. Elaborate and thorough tests have demonstrated that no deterioration of the cables will result from the use of this gas. Some additional precautions are required in entering manholes and using high-temperature torches for soldering when the SF_6 gas might be present.

Alarm Equipment

Since the auxiliary repeaters and certain of the main repeaters are unattended, it is necessary that arrangements be provided to indicate at the attended stations when some piece of equipment fails to perform satisfactorily.

Auxiliary repeaters using pilot regulators are equipped with microammeter relays which monitor the operation of the regulator continuously. These relays provide an indication of the operation of the regulator and the power of the 7,266-kc pilot at the output of the repeater. When conditions change from the nominal by a specified amount, the relay contacts close and are locked magnetically. This bridges an alarm pair in the cable and operates an alarm at the nearest attended repeater. By means of Wheatstone bridge measurements from this repeater over the same alarm pair, the repeater in trouble can be located and a maintenance crew dispatched to make the necessary equipment replacements. The relays can also be reset over the same alarm pair to aid in the location process or to clear alarms which were initiated at unaffected repeaters by deviations in the pilot due to troubles at preceding repeaters.

At main repeaters, microammeter relays are provided on all six pilots used to control the equalization of the system. Deviations in these pilots operate the automatic switching equipment and initiate the usual office alarms. Alarms are also provided to indicate fuse operation, transfers from regular to stand-by equipment, and the condition of electron tubes in the terminal equipment amplifiers.

Provisions for connection to special alarm systems are made at main repeaters which are not fully attended. These systems extend the alarms to the nearest attended repeater and enable the attendant to determine in considerable detail the condition at the remote repeater. The attendant may also perform certain operations such as switching a working line to a spare line at the remote repeater.

Maintenance

Maintenance of the *L3* system requires equipment and methods for routine checking of the system and trouble location. Normally the auxiliary repeaters will be visited at intervals of about 3 months, when checks will be made of the power voltages and currents, the electron tube bias and change in bias with a fixed change in heater voltage (activity), and the pilot magnitudes. At these times amplifiers and regulators which fail to meet prescribed limits will be replaced, the 7,266-kc pilot will be brought to its normal value by adjusting the regulator gain, and the amplifier gain control in the output beta circuit will be adjusted by observing the 3,096-kc pilot.

For these routine tests and adjustments two portable test sets are provided. The power test set plugs into the repeater, amplifier, or regulator and provides for measuring power supply voltage and currents and electron tube cathode-grid voltages to an accuracy of ±1 per cent. The pilot indicator makes it possible to measure the 7,266- and 3,096-kc pilots to an accuracy of ±0.1 decibel.

For trouble locations at auxiliary repeaters a portable transmission measuring set has been designed. It is capable of measuring the power in a 500-cycle band at any place in the frequency spectrum from 50- to 11,000-kc to an accuracy of ±0.5 to ±0.02 decibel, depending on its specific use and the care used in calibration.

At main repeaters, line sections will be checked for noise and modulation performance and equalizers will be adjusted at intervals of 1 week to several months. In addition, loss and gain measurements will be made on sections of the office suspected of being in trouble. For all general tests except equalization line-up, point-by-point measuring equipment is provided, consisting of a 50- to 10,000-kc oscillator, the tuned transmission measuring set referred to above, a milliwatt power meter accurate to ±0.035 decibel and a complement of attenuators, pads, and comparison switches.

The adjustment in a rapid and direct way of the manual gain equalizers to an accuracy of ±0.02 decibel is accomplished by special equipment described in the companion equalization paper.[10] A visual gain and delay transmission measuring test set, capable of measuring gain to ±0.05 decibel and delay to ±0.02 microsecond, has been developed for observing the line performance and adjusting delay equalizers when the system is used for television.

A maintenance center is provided at about 200-mile intervals along the line to service the equipment removed from repeaters. At these points facilities are provided for the following:

1. electron tube testing;
2. regulator repair and adjustment;
3. transmission measurements on passive components; and
4. amplifier testing of sufficient scope to permit changing tubes and to determine whether an amplifier is suitable for further service in the line.

References

1. Progress in Coaxial Telephone and Television Systems, **L. G. Abraham.** *AIEE Transactions*, vol. 67, pt. II, 1948, pp. 1520–27.

2. Frequency Division Techniques for a Coaxial Cable Network, **R. E. Crane, J. T. Dixon, G. H. Huber.** *AIEE Transactions*, vol. 66, 1947, pp. 1451–59.

3. Load Rating Theory for Multichannel Amplifiers, **B. D. Holbrook, J. T. Dixon.** *Bell System Technical Journal*, New York, N. Y., vol. 18, Oct. 1939, pp. 624–44.

4. Cross-Modulation Requirements on Multichannel Amplifiers Below Overload, **W. R. Bennett.** *Bell System Technical Journal*, New York, N. Y., vol. 19, Oct. 1940, pp. 587–610.

5. Data on Random Noise Requirements for Theater Television, **Pierre Mertz.** *Journal*, Society of Motion Picture and Television Engineers, New York, N. Y., vol. 57, Aug. 1951, pp. 89–107.

6. The Subjective Sharpness of Simulated Television Images, **M. W. Baldwin, Jr.** *Bell System Technical Journal*, New York, N. Y., vol. 19, Oct. 1940, pp. 563–86.

7. Single Frequency Video Interference Data, **M. W. Baldwin, Jr.** *Proceedings*, Joint Technical Advisory Committee, Annex 8, vol. 2, Dec. 1948.

8. L3 Coaxial System—Equalization and Regulation, **R. W. Ketchledge, T. R. Finch.** *AIEE Transactions*, vol. 72, pt. I, Sept. 1953, pp. 414–30.

9. L3 Coaxial System—Amplifiers, **L. H. Morris, G. H. Lovell.** *AIEE Transactions*, vol. 72, pt. I, 1953 (*Paper 53-222*).

10. L3 Coaxial System—Television Terminals, **John W. Rieke, R. S. Graham.** *AIEE Transactions*, vol. 72, pt. I, 1953 (*Paper 53-224*).

11. Quality Rating of Television Images, **Pierre Mertz, A. D. Fowler, H. N. Christopher.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 38, Nov. 1950, pp. 1269–83.

12. L3 Coaxial System—Quality Control Requirements, **H. F. Dodge, M. K. Kruger, B. J. Kinsburg.** *Bell System Technical Journal*, New York, N. Y., vol. 32, July 1953, pp. 943–67.

◆

No Discussion

L3 Coaxial System—Equalization and Regulation

R. W. KETCHLEDGE
NONMEMBER AIEE

T. R. FINCH
NONMEMBER AIEE

Synopsis: The equalization and regulation problems of the *L3* system are described and a theory of equalization of complex systems is outlined. The location and function of the various equalizers are explained, including the roles and design of the various fixed, dynamic, and manual equalizer networks. The analogue computer used in the regulation system is described, together with the cosine-equalizer adjusting technique used with manual equalizers. Finally the circuits and operation of the regulation system and its components are presented.

EQUALIZATION is the process of correcting system gain and delay deviations sufficiently to permit the satisfactory transmission of signals. Regulation refers to that part of equalization which, by automatic means, corrects for relatively rapid changes in transmission. In the *L3* coaxial system the transmission of television signals through more than 1,000 amplifiers introduces relatively severe equalization and regulation problems. Compared to the *L1* system, the *L3* system has nearly three times the bandwidth and over three times more stringent transmission objectives. Thus it has been necessary to devote considerable effort toward finding equalization methods that yield practical and economical solutions.

In previous systems it has often been the practice to design the bulk of the equalization system after the completion of an initial installation and the determination of the deviation characteristics of the system. To expedite the introduction of the *L3* system into the field, equalization study and planning were initiated at the very beginning of the development of the system. One of the design methods was to make highly detailed studies using relatively inadequate data to find the major problems. As the data improved, the studies were likewise improved. While on the surface it might appear more practical to have deferred this work until the data are more adequate, it has been found that these speculative studies are the only sure guide to getting the right data. Faced with such a complicated problem, it is difficult to select from the vast amounts of factors that might be important those relatively few items on which success or failure depends. In the *L3* system there have been a number of critical problems which required intensive effort to find acceptable solutions, for example, the design of stable regulators to permit operation of 700 regulators in tandem, the choice of shapes for manual and dynamic equalizers, and the selection of equalizer adjustment methods.

The design of equalization for the *L3* system is not yet complete since the dynamic equalizer shapes are still subject to considerable uncertainty and the details of some of the final television mop-up equalization are yet to be settled. However, no major difficulties are anticipated in equalizing the telephone system to be installed from Philadelphia to Chicago during 1953. Although amplifiers having the final gain characteristic were operated in the trial line between New York and Philadelphia for the first time in November 1952, it was immediately possible to transmit quite satisfactory television pictures over the 200-mile loop using existing equalizers. It therefore appears that equalization will not limit the rate of field installation of the *L3* system.

The Problem

The 4,000-mile coaxial cable has a gain distortion of nearly 40,000 decibels between 0.3 and 8.5 megacycles. Although the amplifiers reduce the distortion to perhaps 200 decibels, and 100 microseconds, they leave a residue characteristic that is considerably more difficult to equalize. Further, it is necessary to deliver service to intermediate offices spaced on the average about 120 miles apart. This requires equalization of high precision at numerous intermediate points. A further problem is the variability of the transmission characteristic due to manufacturing deviations plus time and temperature changes. Also, gain distortion cannot be permitted to exceed about 5 decibels at any point in the line or the signal misalignment will result in degraded signal-to-noise ratios.[1]

The over-all transmission objectives[1] are of the order of 0.25 decibel and 0.1 microsecond which, if allocated among the more than 1,000 amplifiers, lead to rather unrealistic amplifier requirements. In fact, individual amplifiers do not always meet the 4,000-mile over-all requirements. Thus it is the problem of the equalization designer to provide a mop-up system that will permit attainment of the transmission objectives at all service points and at all times.

Equalization Theory

One of the steps in the solution of the general problem has been to develop a theory of equalization. This theory merely applies information concepts to the equalization problem to determine what information is required, when it is needed, and how it may best be used. This theory has stimulated the development of novel equalizer adjustment techniques and has been of assistance as a guide to the attack on the general problem.

To equalize a system, the man (or machine) who is to perform the action must know what corrective steps are required, and for this he must have some kind of information as to the present state of the system and as to the desired state. Second, he must have the necessary tools to convert the system from its present state to the desired state. Consider the first problem, the determination of the corrective steps required. The problem is to determine what we need to know, when we need to know it, and especially in what form the information may be utilized most efficiently.

It can be assumed that the desired state of the system is known, which is usually a constant loss with constant delay over the frequency range of interest. As to the present state of the system, it may be noted that sufficient information can never be obtained to equalize a system perfectly because of the finite bandwidth of the system and because the system changes with time. This is not a new

---

Paper 53-223, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 18, 1953; made available for printing April 24, 1953.

R. W. KETCHLEDGE and T. R. FINCH are with the Bell Telephone Laboratories, Inc., Murray Hill, N. J.

Space does not permit listing all who contributed to the success of this work. However, the authors wish to mention S. A. Levin for his work on cause shapes, R. H. Klie for his leadership in system studies, E. T. Harkless for his study of cosine equalizers, and E. Ley for mechanical design of equalizers. Also, mention should be made of C. J. Custer and E. G. Morton for their work on regulator circuits, A. R. Rienstra for studies of gain enhancement, C. H. Bidwell for chain action analysis, and F. R. Dickinson for mechanical design of regulators.

fact, nor apparently a very important one, because the system need not be perfectly equalized for satisfactory transmission of signals. It leads, however, to the converse idea, which is important, namely that out of this infinite amount of information regarding the state of the system one should collect only the minimum amount needed. This implies making no more measurements of the state of the system than are absolutely necessary to perform the correction to a degree permitting satisfactory transmittal of the signals. The main purpose of this is, of course, to economize on the time and effort required to obtain information, but it should be noted that excess information may be a source of confusion to the equalization operator.

To put this in the form of a rule, we have:

Rule 1. "Collect only that minimum of information as to the state of the system as will permit equalization to the required degree for satisfactory transmission of the signals."

Having established that a minimum of information should be collected, we return to the time variation of the system, which in theory can make the information obsolete before it can be used. However, without yet bringing in the practical fact that its most rapid rate is relatively slow, we should introduce the fact that the ways in which the system can vary at its most rapid rate are quite restricted as compared with the manner in which it can vary at slower rates, and we note that even these are relatively limited. (This probably applies to most transmission systems, not just to the *L3* coaxial.)

Thus, while the information regarding the more rapid but simple changes must be collected more frequently, it consists of a small amount of information per sample; whereas the information regarding the slow, but more complex, changes need not be collected very often, but it represents a relatively large amount of information per sample. Since the total rate of information collection is proportional to the information per sample times the rate of sampling, the minimum information collection principle would say that the rate of sampling should also be held to a minimum.

This demonstrates the value of association of particular types of system change with the rate and amount of their variation, because one may thus eliminate from the more rapid sampling the collection of information about changes that occur at slow rates. Furthermore, one may establish the sampling rates for the various system effects at the lowest possible value. (There is also a very practical value in knowing the rates and amounts of the various deviations, because system misalignment requirements force the equalization to be suitably distributed along the line.)

In the form of a rule, this is:

Rule 2. "To the greatest practicable extent the over-all system behavior should be separated into individual effects, each having its own time rate of occurrence, and corrections should be made for each effect at the minimum tolerable rate for each."

It is of interest to note that, if we couple the logic of rule 2 with the fact that the fastest changes, caused by changes in the temperature of the repeater huts, take hours to become appreciable, we see that the continuous collection of information from continuous pilots (and the continuous correction by pilot-controlled regulators) is in principle unnecessary and inefficient, except of course for its other function of giving alarms under trouble conditions.

The technical problem of determining the equalization states of the system is normally solved by sending some kind of signals over the system and observing the effect of the system on those signals. The raw data are usually in the form of loss and delay as a function of frequency. On the basis of these data, the equalization operator desires to correct the system by means of some equalizers which have adjustable transmissions and delays as a function of frequency. Thus the operator has a group of controls to be operated plus some data which has encoded in it the information as to the proper adjustment of each control. From these data, and a knowledge of the effect of each control, the operator must suitably compute the proper adjustments. As this may be too complicated a process to attempt on on a trial and error basis, or by numerical methods, it is quite an obvious advantage to the operator to receive the data as to the state of the system, not in its original form, but in the form of the necessary adjustments to his equalizer controls. This new form on the data simply represents a decoding process based on the available controls. An operator with the same data, but different and perhaps more complicated equalizers, would need the data in a form suited to his different equalizers.

Consequently:

Rule 3A. "The information as to the state of the system may best be presented to the equalization operator in the form of the necessary adjustments of the available equalization controls."

This rule has a closely related corollary which is based on the fact that the available equalization controls determine the amount of information needed. For example, if the independent gain equalization controls are n in number, measurement of the gain of the system at n suitably chosen frequencies is sufficient to determine the settings. (If the controls are not independent, fewer than n frequencies need be measured.) This is, of course, a restatement of the fact that n unknowns may be determined by solu-

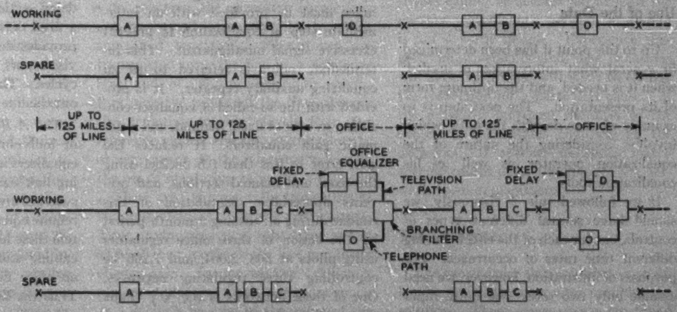


Fig. 1. Typical equalizer locations in telephone systems and in combined systems

tion of n independent simultaneous equations. The unknowns are the equalizer settings and the simultaneous equations are the relationships of the shapes controlled by each equalizer to the total system error.

Thus:

Rule 3B. "In general, the necessary and sufficient condition for the determination of n independent equalization control settings is the knowledge of the system's equalization error at n independent frequencies plus the knowledge of the effect of each of the n controls at each of the n frequencies."

It should be noted that this statement of the rule assumes analysis by frequency rather than by transient behavior. This approach is used because, at present, equalizers are usually designed on a frequency characteristic basis. The validity of the rule is, however, more general.

From these two rules we can derive another regarding the minimum information principle, which is similar to rule 1 but is actually quite independent of it.

Rule 4. "No more information should be gathered from the system than is necessary to provide sufficiently accurate control setting information for the equalization operator."

In this case the superfluous information may actually cause confusion or harm. It will, at the very least, confuse a manual operator to know of an error he is powerless to correct, whereas a mechanized system would probably go wild if it obtained too much information.

It is evident that the minimum amount of information that would be obtained in accordance with rule 4 would be the same as that obtained in accordance with rule 1 only if the design of the equalization were optimum, because then the equalizer shapes, and the number of shapes, would just suffice to permit satisfactory transmission of the signals.

Use of the Data

Up to this point it has been determined in general what information is needed, when it is needed, and the optimum form of its presentation. The next step is to examine what to do with the data, which involves considering the nature of the equalization operator as well as his equalization tools.

If we followed rule 2 vigorously, we should have several different types of controls, one for each of the effects having different time rates of occurrence. For purposes of illustration, however, we need assume only two rates, one quite rapid and the other very slow. It will be postulated here that very rapid equalization operations are most economically performed by machine such as, for example, the automatic regulation for cable temperature variations. Likewise, relatively infrequent adjustments will be assumed to be best performed by a suitably informed human operator.

The first principle to note is that the only real distinction here is the rate at which data should be refreshed and acted upon. In either case the data should be in the same form, a set of numbers (or their equivalent) representing equalizer changes.

Probably the most useful result of this theory of equalization has been the conclusion that mechanization of the equalization process, particularly with regard to computational techniques, permits substitution of simple logical methods for inefficient trial-and-error adjustment processes. This led to the use of an analogue computer in the regulation system and, for manual equalizers, the development of measuring circuits that read directly in terms of equalizer adjustment error.

Equalization

LOCATION AND FUNCTION OF EQUALIZERS

Fig. 1(A) shows the location of equalizers in the *L3* system when only telephone is transmitted. Combined telephone-television equalization is shown in Fig. 1(B). When telephone and combined systems use the same spare line, that line is equipped for television. The general features of the main-repeater layout have been described.[1] The detailed layout of equalization is designed to meet the requirements of both telephone and television service and the need for flexibility in television network arrangements. In addition, switching of telephone or television service to a spare line must not appreciably degrade service.

Switching sections longer than 120 miles must be provided with an intermediate step of equalization to prevent excessive signal misalignment. This intermediate point is referred to as an equalizing auxiliary repeater. It is provided with the so-called *A* equalizer consisting in turn of fixed, manual, and automatic gain equalizers. It reduces the gain error to less than 0.5 decibel using the fixed and manual sections, and prevents appreciable degradation of this residue during an ensuing 3-month period by the action of three office regulators using pilots at 308, 2,064, and 7,266 kc controlling three regulating networks. One of these networks is the $\sqrt{f}$ shape of the receiving amplifier.[2] The other two are first order corrections for vacuum-tube aging and repeater temperature changes.

At offices (switching main repeaters) on telephone systems further equalization is required to permit line switching of telephone channels and special service signals, such as telegraph. Also, it is not practical to provide telephone equalization on a cumulative basis over longer links than a switching link because of frogging and dropping.[1] No telephone channel rides for more than 800 miles in the same frequency location, and dropping breaks up the pattern still further. Thus, for telephone, the switching links, which may number between 30 and 40 in a long system, are independently equalized.

The *B* equalizer contains manual and automatic sections, the latter being three regulating networks omitted in the *A* equalizer. Thus an *A* plus a *B* forms the final telephone equalization. The residue of five independently adjusted *AB* links must meet telephone requirements without frogging, and 30 to 40 links must meet these requirements with 800-mile frogging. This performance must continue to be met in the presence of a normal amount of spare-line switching. Also it is very important that the character of the residues be such as not to throw an undue burden on the television equalization.

In combined systems the more stringent requirements require the addition of further manual equalization to the switching section. The residue at the output of a *C* equalized switching section must be sufficiently small that a 4,000-mile circuit will continue to meet television transmission requirements in spite of a normal amount of spare line switching. Also the *C* equalizer in conjunction with the *AB* must permit several switching links to be connected in tandem without further line equalization. The *C* equalization also contains an adjustable delay section which in conjunction with a fixed delay equalizer in the office path provides delay equalization for the television part of the band, 3.6 to 8.5 megacycles. This adjustable section builds out the line to match the fixed unit.

The *A-B-C* pattern is for equalization of individual switching lines and these equalizers are adjusted on a single switching link basis. On the office side of the switches are system components that also require equalization. In the telephone system these include such equipment as office cabling and hybrid coils. Also there is an office flat loss of nearly 30 decibels. Thus the lines are, in effect, operated to give a 30-decibel gain while the offices give a 30-decibel loss. Aside from the flat losses there are distortion shapes but fortunately these are all well approximated by a $\sqrt{f}$ shape and a single manually adjustable $\sqrt{f}$ equalizer plus suitable choice of flat loss provides adequate telephone office equalization. This is referred to as the 0 equalizer. It is adjusted to cause the particular office to have a flat characteristic.

In the office circuits of combined systems are branching filters to separate the telephone and television bands. The 0 equalizer is reused in the telephone path. The television path includes the fixed delay equalizer and a manual gain equalizer to correct for office cables, hybrids, etc. After the tandem combination of several independently equalized lines and offices further equalization will be required. This will be accomplished by a multicontrol manual *D* equalizer having both gain and delay sections, inserted in the television-only path at approximately 400-mile intervals. These *D* equalizers will be used to form 800-mile pilot links which are independently equalized to a degree permitting putting any five such links in tandem to form a 4,000-mile circuit without further equalization.

Fixed Equalizers

The line amplifier can properly be considered as the first step of fixed equalization.[2] Also acting at this same level are artificial cable networks used to build out the repeater spacing to 4 miles±0.2 mile, as well as the basic equalizer of the amplifier to take up differences between cable types.[1] The final step of fixed equalization is the so-called design deviation equalizer associated with all *A* equalizers. This equalizer comes in two forms, one for use with sections containing 23 to 32 repeaters and one for use in sections of 10 to 22 repeaters. In those few cases of less than 10 repeaters the fixed equalizer is omitted.

The function of the design deviation equalizer is, first, to correct for the design error of the average repeater and, second, to recenter the manual (cosine) equalizers. Although the average repeater matches its 4 miles of cable to within ±0.12-decibels, this design error accumulates to over 3 decibels in 30 repeaters and, further, the shape is a difficult one to equalize. To keep the number of designs to a minimum, only two sizes are used, 19- and 28-repeaters, and the residue is corrected by the manual equalizers. This residue may be positive or negative depending on whether the fixed equalizer over- or under-compensates. Thus there is only a small tendency for these residues to accumulate in long systems. However, the inaccuracies of match between the fixed equalizer and the gain of the average repeater section tend to accumulate systematically and must therefore be kept small. This brings out the importance of statistical quality control of amplifier manufacture,[2] since any systematic shift in the amplifier gain characteristic will accumulate and may consume excessive manual equalizer range or may lead to excessive equalization errors. For example, a shift of only 0.05 decibel in the gain of the average amplifier would represent 1.5 decibels in 30 repeaters and 50.0 decibels in 1,000 amplifiers, and thereby becomes an extremely serious matter. Thus quality control of the amplifier is a vital part of the solving of the equalization problem.

The various effects that consume the range of the manual equalizers produce for some shapes an unsymmetrical consumption of range. If uncorrected this would produce larger range requirements in the manual equalizers as well as introducing new shapes to be equalized. The manual equalizer shapes are symmetrical and their errors cancel if equal amounts of positive and negative range occur in the system. Any systematic offset of a particular shape tends to introduce new shapes because of the manual equalizer networks themselves. By appropriate modification of the shape of the fixed equalizer it is possible to recenter the manual equalizers so that on the average the manual shapes are in the center of their range.

Dynamic Equalizers

GENERAL

Any long transmission system suffers from relatively rapid gain changes and in the *L3* coaxial system, as in many previous systems, the necessary corrections are performed automatically by pilot controlled regulators. Pilot tones are transmitted over the line at a reference level and, at appropriate points, regulators pick the pilots off the line, observe the deviation in pilot levels from the reference values, and restore the pilots to or very nearly to the reference values by the use of regulating networks.

There are fundamentally two causes of fast gain changes, time and temperature. Time produces vacuum tube aging and in spite of their feedback the line amplifiers change gain. In one week a 4,000-mile system is expected to change by as much as 5 decibels because of the aging of the 6,000 tubes or so in the transmission path. Because of different thermal characteristics, temperature affects cable and repeaters semiindependently. In a 4,000-mile system 1 week is expected to eugender as much as an 8-decibel gain change by change in repeater temperature. Cable changes can exceed 100 decibels per week. Thus these effects must be corrected to a very high order of precision to maintain good television or telephone service.

These gain changes are not the entire story; when the gain changes in the band it also changes outside the band and usually by an even larger amount. Thus equalization of the in-band gain leaves out-band (above 8.5 megacycles) gain changes which produce in-band delay changes of several microseconds. Because of the difficulty of providing automatic delay equalization, a complicated process, it is necessary to equalize these delay changes on a gain basis, by at least partial correction for out-band gain changes. Further, since satisfactory pilot transmission is possible only within the band, these out-band changes must be predicted from the in-band gain changes. This effect in itself indicates the use of equalizers whose individual shapes are those produced by specific system causes producing a correlated change in many elements. Thus building the regulating networks to match the effects of the individual system causes, and matching to 10 or 12 megacycles rather than merely to 8.5 megacycles, permits simultaneous gain and delay equalization. Such a set of shapes is also more accurate because it is matched to the special ways in which the specific system can change rapidly.

In theory, the regulating networks could be built to match linear combinations of the cause shapes but there are two difficulties. First, not all of the cause shapes are known accurately or often even roughly at the introduction of a new system into the field. The mixtures of shapes cannot be determined without the missing ingredients. Second, a system is not a static design. Experience suggests improvements and thus occasions will arise where one will want to change the correction for a particular cause. If the networks represent mixtures of the causes, it necessitates changing all the networks. If specific networks match specific causes only, the appropriate network needs replacement.

The Computer

The use of cause shapes leads to a problem to which the computer provides the solution. These cause shapes are broad effects covering the entire band and more. Thus no one pilot is a measure of a specific cause. However, by a process equivalent to the solution of simultaneous equations, the pilots determine the amounts of an equal number of cause shapes that will restore the pilots to normal. Thus the computer translates pilot errors into shape errors and drives the appropriate regulating networks to obtain the corrections.

Let the equalizer shapes be given by functions of the form

$$S_n(f)=k_nF_n(f) \tag{1}$$

where

n = the subscript number identifying the particular equalizer. The capital N is reserved for the total number of equalizers.
$F_n(f)$ = the equalizer shape (on a unit basis) as a function of the frequency f
k_n = the amount of shape introduced by adjustment, and may be positive or negative
$S_n(f)$ = the resultant shape put in the system by adjusting $F_n(f)$ by an amount k_n

The total shape introduced by all N equalizers is obviously

$$S_{\text{total}}(f)=\sum_{n=1}^{n=N} S_n(f)=\sum_{n=1}^{n=N} k_nF_n(f) \tag{2}$$

To obtain a match of S_{total} to the given equalization error, S_{given}, at M frequencies from $m=1$ to $m=M$, requires that

$$S_{\text{total}}(f_m)=S_{\text{given}}(f_m) \tag{3}$$

at each frequency from f_1 to f_M. Or, in terms of equation 2

$$S_{\text{given}}(f_m)=\sum_{n=1}^{n=N} k_nF_n(f_m) \tag{4}$$

again, at each frequency from f_1 to f_M.

All of the important conclusions regarding the action of an equalization computer are implicit in the M equations indicated by equation 4.

Consider a case where there are three shapes. Let the information as to the difference between the system state and its desired state be determined by the deviation of three pilot levels which are observed to be δ_1, δ_2 and δ_3 at the pilot frequencies f_1, f_2 and f_3. The problem is to find the values of k_1, k_2 and k_3 that will give a match at these frequencies. This means that the following equations must be satisfied

$$S_1(f_1)+S_2(f_1)+S_3(f_1)=\delta_1 \tag{5}$$

$$S_1(f_2)+S_2(f_2)+S_3(f_2)=\delta_2 \tag{6}$$

$$S_1(f_3)+S_2(f_3)+S_3(f_3)=\delta_3 \tag{7}$$

Thus

$$k_1F_1(f_1)+k_2F_2(f_1)+k_3F_3(f_1)=\delta_1 \tag{8}$$

$$k_1F_1(f_2)+k_2F_2(f_2)+k_3F_3(f_2)=\delta_2 \tag{9}$$

$$k_1F_1(f_3)+k_2F_2(f_3)+k_3F_3(f_3)=\delta_3 \tag{10}$$

The solutions for k_1, k_2 and k_3 take the form

$$k_m=a\delta_1+b\delta_2+c\delta_3 \tag{11}$$

where a, b, and c depend solely on the shapes $F_m(f)$.

Thus the values of the δ's may be decoded into the values of the equivalent k's by the simple process of multiplying each δ by some fraction that is a function of the equalizer shapes or, more precisely, by a fraction that is a function of the values of the various shapes at the frequencies f_1, f_2, etc.

The circuit of the computer is quite simple, consisting of about N^2 resistors for the control of N networks by the deviations that are measured at $M=N$ pilot frequencies. This simplicity is valuable for its own sake but, as previously noted, there is considerable value in the fact that substitution of new equalizer shapes requires only that changes be made in the values of some of the resistors.

Fig. 2 illustrates the principle by showing the computer circuit required for the previous example of three shapes for which information is given at three frequencies, f_1, f_2 and f_3. The 3 direct voltages representing the deviations δ_1, δ_2 and δ_3 are decoded by simply cross-connecting them to the three pairs of output terminals through fixed resistors chosen to satisfy the relationships of equations 8, 9, and 10. The output voltages will be proportional to the desired equalizer correction quantities k_1, k_2 and k_3.

In general, the calculation of some of these resistors will give negative values. Thus it will generally be necessary to require that the direct voltages representing the errors δ_1, δ_2, etc., be available in both polarities. Alternately, the errors can be provided in only one polarity, and the circuits to which the computer outputs connect can provide the push-pull circuit. This latter course has been used in the L3 regulators as indicated in Fig. 2.

The effect of using the computer is as follows. If one pilot changes, all regulating networks correct, but in such proportions and polarities as to produce no gain change at any pilot frequency except at the one originally disturbed. If all pilots deviate in proportions corresponding to one of the shapes, no network corrects except the one corresponding to the original pattern of pilot deviation.

If conventional regulator circuits were used with particular pilots assigned to

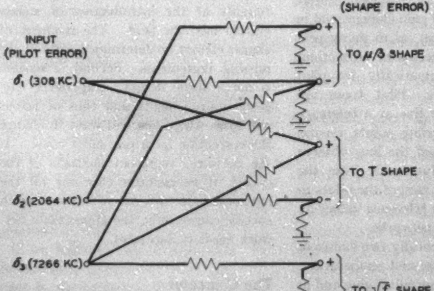


Fig. 2. Schematic of regulation system computer

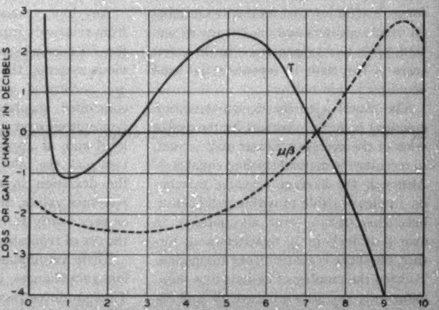


Fig. 3. Regulation shapes and ranges for tube aging $\mu\beta$ and repeater temperature T. The curves given can occur as either gain or loss and apply to 30 repeaters

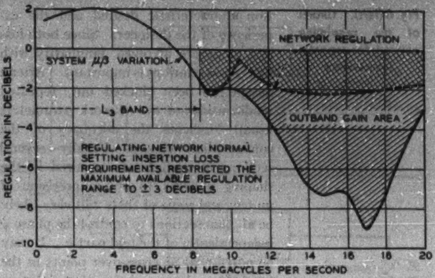


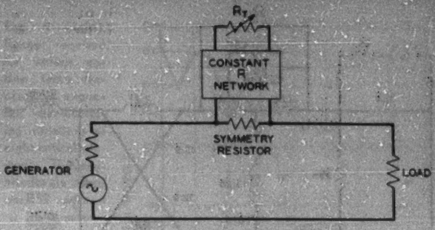
Fig. 4. Match to out-band gain change to reduce in-band delay distortion

particular shapes, the interactions would be intolerable. Thus the computer removes the restrictions on the choice of shapes imposed by regulator interactions. In turn, this permits freedom in changing shapes as system data improve. It is important to provide the initial equalizers before adequate data on the system are available. As the system grows in length, the equalization must be improved but the data improve also. Thus there is an economic benefit in providing a flexible equalization plan which allows the earliest possible commercial use of the system.

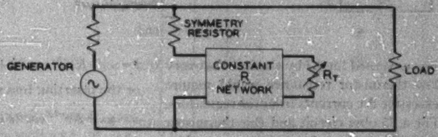
DYNAMIC EQUALIZER REQUIREMENTS

Because temperature and aging variations can result in both gain and loss variation with respect to the average transmission, the equalizers must be capable of inserting both loss and gain compensation. Also, it is extremely desirable that equal gain and loss settings for the equalizer result in symmetrical transmission characteristics with respect to the average. In a long transmission system, some of the sections will insert excess gain and others, excess loss. If the equalizer gain and loss characteristics are symmetrical, the residue will be related to the system deviation and some constant part of the equalizer characteristics. If this is not the case, the residues can produce new variable deviation shapes and thereby lead to increased complexity in following the stages of equalization.

As previously described, equalization of both gain and loss is required and this implies active equalizers. Although a number of methods were studied, noise and modulation requirements led the *L3* system to the so-called block-of-gain–block-of-loss design. The equalizers are passive networks and, to realize both gain and loss adjustability, the normal setting loss must at least equal the total amount of gain adjustment. The various equalizers are combined into two to four groups whose loss is compensated by corresponding numbers of flat gain amplifiers.[2]



(a) SERIES TYPE



(b) SHUNT TYPE

Fig. 5 (right). Block diagram of Bode regulating network

The required number of such blocks of loss and gain depends upon the amount of system gain variation to be equalized. The determination of the shapes and magnitudes of system variations is an important system problem. Extensive study is necessary to evaluate the system sensitivity to various changes; the determination of the magnitude of these causes, such as temperature variation, aging rates, etc., and the determination of maintenance intervals that provide an economic balance between maintenance expense and system cost. These must be studied in detail to provide the equalizer designer with shape and range data for his dynamic equalizer designs. The equalization characteristics and maximum ranges for the two most important dynamic equalizers aside from cable temperature, namely, repeater temperature T and vacuum tube aging $\mu\beta$ are shown in Fig. 3. To maintain satisfactory transmission during a maintenance interval, the dynamic equalizers must be able to match any characteristic within the maximum ranges and throughout the transmission band to within 1 to 2 per cent. As noted previously, the necessity for simultaneous delay equalization requires the dynamic networks also to make at least an approximate correction in the out-band region. This is shown in Fig. 4.

The control element in the equalizers is a thermistor whose available resistance range is 30 to 1,050 ohms. ($\sqrt{f}$ of line amplifier,[2] 125 to 2,000 ohms). Thus the regulation ranges shown in Fig. 3 are realized, using this one variable resistance element in each equalizer.

The following is a summary of the dynamic equalizer requirements:

1. To provide symmetry in regulation characteristic.
2. To minimize flat loss.
3. To match prescribed gain variation to a high degree of accuracy within transmission band.
4. To provide satisfactory equalization of in-band delay distortion caused by out-band gain change.
5. Each equalizer to be controlled by a single variable resistance element.
6. The desired performance to be realized in 75-ohm circuits.

DYNAMIC EQUALIZER DESIGN

The design used is the structure commonly known as the Bode Regulator,[3,4] and shown in block schematic form in Fig. 5. An ideal regulating network would insert a deviation characteristic

$$\theta = f(R_T)f(\omega) \text{ decibels} \qquad (12)$$

Such a network would yield a given shape independent of setting and thus would have linearity as well as symmetry. Actually the Bode network has symmetry but only approximates linearity. As shown, it may be designed to insert a variable impedance either in series or in shunt with the line. When using only one control element it is not a constant resistance structure and must be operated between terminations of the design value. The choice of series or shunt configuration

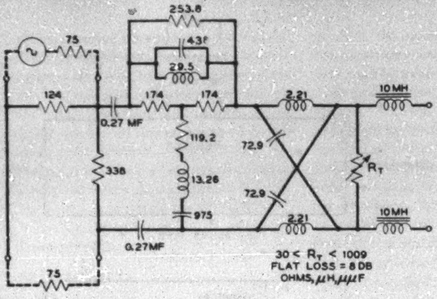


Fig. 6 (left). Circuit of the regulating network for repeater temperature

was governed for *L3* by the fact that very low thermistor resistance would require excessive d-c currents from the regulators. For a 75-ohm circuit and this thermistor limitation the series regulator provides lower normal-setting loss.

The insertion gain for the series regulator may be written as

$$\theta=\theta_N+2\,\text{arc tanh}\left[\frac{R_1-R}{R_1+R}\rho\epsilon^{-2\Gamma}\right] \tag{13}$$

where

θ_N = the normal setting flat loss as given by

$$\alpha_N=20\log\frac{R_1}{R}\ \text{decibels} \tag{14}$$

R_1 is the symmetry-determining resistor and is in parallel with the input terminals of the shaping network for the purpose of limiting the maximum value of the series impedance. The equation of symmetry may be expressed as

$$R_1^2=\left(1+\frac{R_1}{2R_0}\right)R^2 \tag{15}$$

where

R = the image impedance of the shaping network. The thermistor has this value at its normal setting.

R_0 = the impedance of the circuit in which the regulator is inserted, in this case, 75 ohms.

ρ = the reflection factor between the network impedance R and the thermistor impedance, R_T, and is given by

$$\rho=\frac{R_T-R}{R_T+R} \tag{16}$$

Γ is the image transfer constant of the shaping network and may be expressed as

$$\Gamma=\sigma+i\psi \tag{17}$$

Returning to equation 13 for the insertion gain, if the series expansion of the arc tangent is used and the higher order terms discarded, the insertion transfer constant becomes

$$\theta=\alpha_N+K\epsilon^{-2\sigma}\cos 2\psi+iK\epsilon^{-2\sigma}\sin 2\psi \tag{18}$$

or the insertion loss may be expressed

$$\alpha=\alpha_N+Ke^{-2\sigma}\cos 2\psi\ \text{nepers} \tag{19}$$

$$K=2\frac{R_1-R}{R_1+R}\rho \tag{20}$$

This approximation is valid for design purposes if the regulation range is not too large or the accuracy required too great. For example, omission of the cubed term in the repeater temperature design produced a maximum error of 0.014 decibel and in the tube aging design, 0.007 decibel.

From this expression it is seen that both real and imaginary parts of the image transfer constant for the shaping network enter into the insertion loss expression, and thus the relationship between the desired insertion loss and the shaping network makes the design process difficult. At this stage in the design considerable art and ingenuity are required to continue the design efficiently. The approach taken depends upon the regulation characteristic desired and the experience of the designer. Since both loss and phase are involved, familiarity with loss-phase relations is important. After a design has been blocked out, repeated modifications of the network parameters usually indicate the adequacy of the configuration chosen. Usually several sections in tandem are required in the shaping network to obtain precision of match, and some of these will very likely be all pass sections to control the phase ψ independently of the real part σ. It may be noted that at cross-over points in the regulation characteristic, that is, zero regulation, either the phase ψ must be 45 degrees, or odd multiples thereof, or the loss σ must be infinite. Usually the phase is made the controlling term in both zero and peak regulation points.

The designs presently used on the *L3* system for the repeater temperature- and tube-aging equalizer are shown in schematic form in Figs. 6 and 7. The 0.27-microfarad capacitors and the 10-millihenry inductors are employed to supply the d-c heating current to the thermistor. The shunt resistors placed across the 75-ohm line, in one case 124 ohms, and in the other, 133.4 ohms, are used to make the left side driving point impedance equal 75 ohms at normal setting of the thermistor. These networks are used in cascade with the manually adjusted, constant-resistance networks, and it seemed desirable that the impedance characteristic of the dynamic equalizers be relatively good at one pair of terminals.

Manual Gain Equalizers

One of the more difficult equalization design choices is the selection of shapes

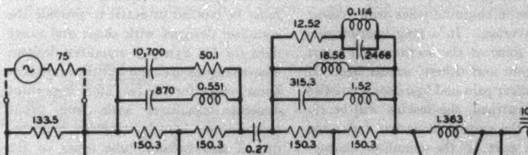


Fig. 7. Circuit of the regulating network for tube aging

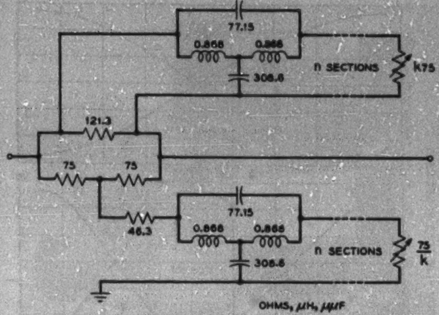
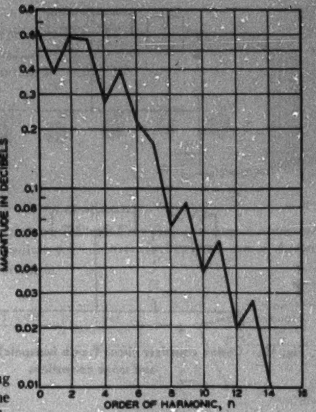


Fig. 8. Circuit of a cosine equalizer without dissipation correction

Fig. 9 (right). Range required for various cosine terms because of expected manufacturing variations of eight critical elements. The magnitudes are based on RSS addition in 25 repeaters



for the manual equalizers. In the *L1* system some of the shapes used were so-called bump shapes. This type of shape reduces the amount of shape overlap and thereby tends to reduce the adjustment problem when conventional adjustment methods are used. For the *L3* system many types of shapes and circuits were carefully studied. The computational concepts of equalizer adjustment resulted in the consideration of shapes that would otherwise be impractical.

The shapes finally chosen for *L3* are cosine shapes. Any continuous function may be matched over a 180-degree interval by a Fourier series of cosines only. By making the 180-degree interval the frequency range from 0 to 8.5 megacycles, the cosines are those of frequency and can match any gain characteristic if enough terms are used. These cosine equalizers have the following advantages:

1. The range required for any term is less than the total shape to be matched, usually less than half.
2. The residues are always higher harmonics and therefore easily matched if necessary.
3. The controls are easily adjusted, using methods to be described.
4. If better equalization is required at any point, the existing equalizer is retained and additional harmonic terms are added.
5. The major portions of the equalizers require only one value of inductance and two values of capacitance to form the delay lines used in the networks. This also assists in the application of distribution requirements.
6. The manual equalization can be designed with a minimum of information about the system characteristics.
7. The equalization is on a least-square error basis rather than minimum peak error.

The networks used to realize these cosine shapes are constant resistance Bode regulating networks[3] employing second-degree all-pass sections.[5] If the phase of the all-pass sections were made proportional to frequency, the transmission performance within the frequency band of interest would be that provided by a Fourier series composed of cosine terms in the variable ω. By appropriate choice of the all-pass sections, the frequency-phase relationship can be warped to give greater weighting to a specified portion of the frequency range. The phase of the all-pass section is given by

$$\psi = 2\cot^{-1}\left[\frac{b}{2}\left(\frac{fc}{f} - \frac{f}{fc}\right)\right] \tag{21}$$

Small b and high f_c weights the low frequencies, the linear-phase case $b=1.2$, $f_c = 10.2$ megacycles gives uniform weighting and large b weights the high frequencies. A $b=2$, $f_c=13.75$ megacycles, was selected for the *L3* equalizers because it weights somewhat the higher frequencies where the television signal is transmitted and, also, each unbalanced bridge *T* network section can be constructed with only four elements, two like inductors and two capacitors. For b's smaller than 2, coupling between the two like coils is required, and for b's larger than 2, an additional element is required.

The all-pass networks are designed on a 75-ohm impedance level and thus the flat, normal setting of each regulating network is 4.18 decibels maximum. The 75-ohm level makes the series and shunt networks identical and also facilitates manufacturing testing. A special dual variable resistor is used as the control element. It has a resistance range of 15 to 375 ohms and provides a regulation range of ±2.78 decibels maximum.

A schematic of this network is shown in Fig. 8. In the case of the O harmonic, flat gain, the phase sections are omitted. For the nth harmonic term, n sections are used in both the series and shunt arms. The constant resistance structure permits the complete equalizer to be formed by a cascade of such networks without interaction effects. This provides a loss characteristic given by

$$\text{Loss} = k_0 + k_1\cos 2\psi + k_2\cos 4\psi + k_3\cos 6\psi + \ldots \tag{22}$$

where ψ is the phase of the individual all-pass section which goes from 0 to 90 degrees between 0 and 8.5 megacycles. The k's are adjusted by means of the dual adjustable resistors. Note that each term of the series is represented by a corresponding equalizer.

Application of this mop-up polynomial to a larger number of *L3* deviation characteristics indicates that considerably less range than the maximum ±2.78 decibels will be required for most of the cosine terms. Fig. 9 illustrates this convergence of the series for the eight largest amplifier manufacturing variations. This and other studies show that after the first three harmonics the range may be reduced. It can be shown,[6] for example, that if the system gain deviations are finite within the range of interest, or interval of convergence, the coefficients of the approximating polynomial will decrease in magnitude linearly proportional to the number of the terms, that is, the nth coefficient will be smaller than some constant divided by n. If the deviation characteristics are continuous and hence have finite first derivatives, then the coefficients of the approximating polynomial will decrease as the square of the number of the term. If all derivatives are finite the coefficients will decrease exponentially. This last case is

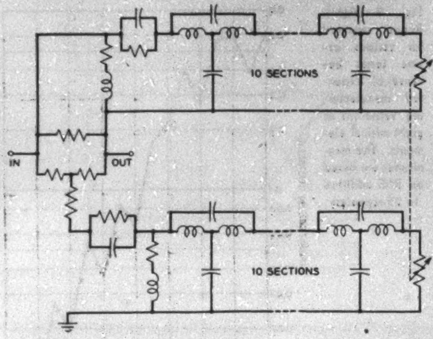


Fig. 10. Cosine equalizer circuit (tenth harmonic) showing dissipation and range corrections

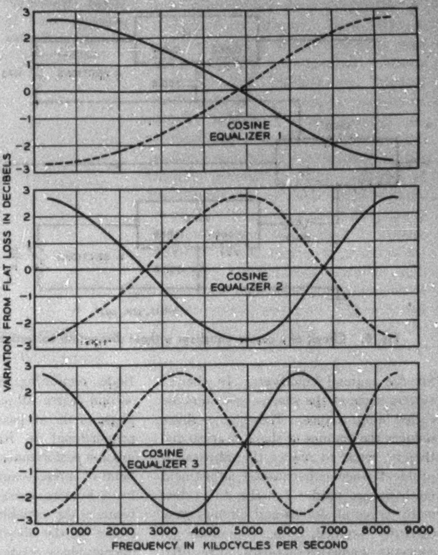


Fig. 11 (right). Shapes introduced by the first three cosine harmonics

believed to describe the convergence for at least most of the *L3* system shapes. Thus for the high order terms that require little range it is possible to reduce the flat loss of the networks.

To reduce the flat loss without changes in the 75-ohm impedance level of the phase sections or in the dual adjustable resistors, pads are inserted between the resistance *T* and the all-pass sections. In this manner the loss could be reduced to 2.2 decibels for ±0.5-decibel range if it were not for the dissipation in the all-pass sections. This dissipation occurs because of the coils, and it increases with frequency. It tends to produce a reduced cosine amplitude in the high-frequency part of the band. To correct this effect, the pad is actually made an equalizer section whose loss change with frequency corrects for the dissipation in the coils, thereby yielding cosine amplitudes independent of frequency. The price of this is an increase in the flat loss to 3.4 decibels for a ±0.5-decibel range. The circuit of the term 10 network is shown in Fig. 10. The range of this term is ±0.5 decibel as is that of all higher terms (to 24). Terms 1, 2, 3 have full range, 4, 5, 6 have 1.5 decibels and 7, 8, 9 have 1 decibel. The 0 term, flat gain, is also 1 decibel because additional flat shape is obtainable from the flat amplifiers used to make up for the equalizer loss.

The shapes provided by the first three terms are shown in Fig. 11. Note that the shapes are cosines of a warped frequency variable and that on this warped scale the shapes are orthogonal. In all, 24 such harmonics plus flat gain are used in the high-frequency line for combined systems. For all-telephone use only 14 harmonics plus flat gain are required. Fig. 12 shows the construction of one of the cosine networks. These are mounted in groups of five as indicated in Fig. 13. The fixed equalizer and regulating networks are mounted to the rear of the cosine assembly.

Harmonic Adjusting Set

To make a mental harmonic analysis of a complicated gain characteristic is difficult, if not impossible. Therefore a special cosine-equalizer adjusting set has been developed which eliminates trial and error from the adjustment process and which leads to a unique optimum adjustment. Broadly, the method consists of using sweep-frequency methods to convert the gain-frequency characteristic into a repetitive voltage-time function. Gain cosines on the warped frequency scale are converted to voltage cosines of time and the audio harmonic-spectrum components are individual measures of equalizer control-setting errors. By the adjustment process these audio harmonics are removed, thus yielding a gain characteristic describable in terms of only the higher cosine components not available to the equalization operator.

The operation can be explained using the block diagram shown in Fig. 14. The sweep oscillator sends a constant-level, variable frequency, over the line and through the cosine equalizer to the detector. The output of the detector on terminals *X—X* at any instant is a measure of the transmission of the line and equalizer at the frequency being sent by the sweep oscillator at that instant. The sweep frequency starts at zero and sweeps to 8.5 megacycles in a period t_1. As shown in Fig. 15 the frequency-time relationship is warped to correct for the warping of the equalizer phase-frequency relationship. The sweep oscillator therefore scans at a linear rate in cosine degrees versus time. Upon reaching 8.5 megacycles the sweep reverses and returns to zero. If the cosine shapes were linear cosines of frequency the sweep would be a triangular wave.

Assume that the line is perfectly equalized except that the first harmonic term is misadjusted. As the sweep goes from 0 to 8.5 megacycles, the voltage at *X—X* follows the first harmonic curve of Fig. 16 from 0 to t_1. When the oscillator scans back to 0, the voltage at *X—X* follows the first harmonic curve from t_1 to $2t_1$. Then the cycle repeats. The voltage at *X—X* thus becomes a pure cosine of time oscillation of frequency, $1/2t_1$. Although the equalizer shapes are actually cosine on a decibel basis rather than an amplitude basis, this has little practical effect because for small deviations the two are nearly identical.

If, instead of a first harmonic error, the second or third cosine harmonic is misadjusted, the voltage at *X—X* will

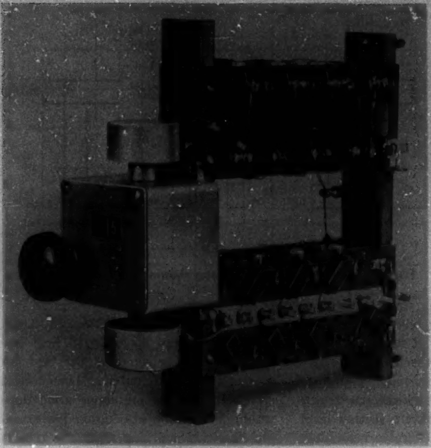


Fig. 12. View of a single cosine term network

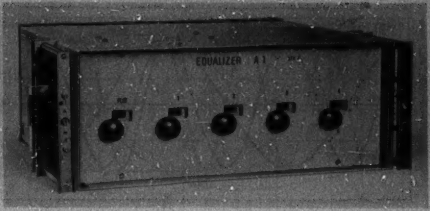


Fig. 13. Assembly of five cosine terms. Five of these assemblies are required by a C equalizer, a B equalizer requires three, and an A, two

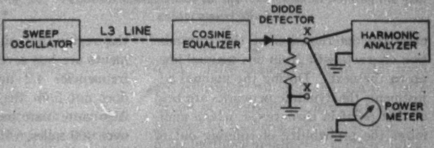


Fig. 14. Block diagram showing the method used to adjust cosine of other orthogonal equalizers

follow the appropriate curve of Fig. 16. The d-c component measures the zero harmonic but in practice only the a-c components are measured and the flat gain is set as a final step to make the pilot levels correct at the line output. If it takes 0.01 second for the oscillator to scan up and back, the first harmonic produces 100 cycles per second (cps) output from the detector. The second harmonic equalizer produces 200 cps, the third 300 cps, etc. Therefore at the output of the detector there exists a set of audio-frequency harmonics whose amplitudes are a measure of the equalization error of the setting of the cosine controls of corresponding periodicity.

These harmonics can be separated by convention filtering techniques, for example, by an audio-tuned detector or harmonic analyzer as noted on Fig. 14. The analyzer can be tuned to 100 cps and the first harmonic control rotated to remove the 100-cps component. Then, tuning to 200 cps, the second control is operated, etc. This null method is similar to bridge balancing and may be instrumented to similar high precision. After all of the harmonics corresponding to equalizer controls have been removed, the process is complete and the equalization residue must be composed solely of those terms not provided by the equalizer.

The harmonic analyzer method requires tuning or switching and thus, to simplify the method still further, the actual field equipment uses a power indicator in place of the analyzer as noted on Fig. 14. Given a spectrum of signals of differing frequencies, removing any one of these reduces the total power. Therefore the entire spectrum may be applied to a power indicator and the reading reduced by adjusting the various equalizer controls. In practice this process is assisted by filtering out the high harmonics which cannot be equalized. This reduces the total power and increases the ease of reading the meter. While the method has been demonstrated, with the use of an ordinary 60-cps wattmeter, an electronic wattmeter is used in the field equipment.

While the receiver unit can be quite simple, the sweep oscillator is complicated by such things as circuits to control the warping and to hold the sweep limits accurately In practice the sweep is between 0.3 and 8.5 megacycles and is at a 37-cps rate. Further there are six pilots on the system and, although the dynamic regulators at the adjusting point are paralyzed during the cosine adjustment, the pilots to intermediate regulators must not be disturbed. Thus the sweep frequency is shifted very rapidly through the pilots. When the sweep frequency gets within about 25 kc of a pilot, it is shifted suddenly to the other side of the pilot frequency. This materially reduces the interference to the pilot without producing transients in the receiver.

While other methods of cosine equalizer adjustment were tested and found to work satisfactorily, the method described was found to be superior. In addition, removal of the filtering permits the power meter to read the rms equalization error and thus the equalization operator can determine the quality of the job and observe whether the state of the line is satisfactory. Also, the power method is usable with saw-tooth as well as triangular scanning and, further, works on any set of orthogonal gain or delay shapes. Thus the basic equipment is readily adaptable to the adjustment of *D* equalizers if their gain and delay shapes are orthogonal.

Field Performance

In the *L3* system the function of the dynamic equalizers is solely to prevent

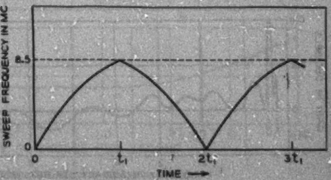


Fig. 15. Method of scanning the system gain characteristic to convert cosines of frequency into cosines of time

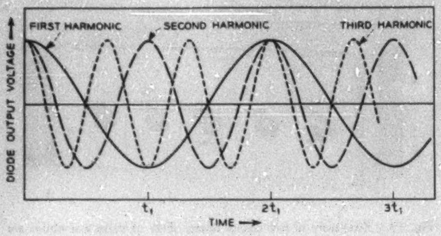


Fig. 16. Detector output produced by scanning the first three cosine shapes

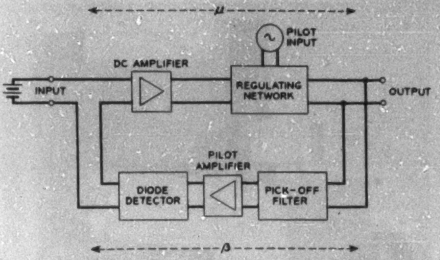


Fig. 18. Block diagram of regulator showing the feedback structure

excessive deterioration of the transmission characteristic from one manual line-up to the next. During the manual adjustment, the dynamic networks are held at that point in their range, which minimizes the probability of running out of range in either direction before the next manual line-up. This is done to hold the dynamic ranges to a minimum. Thus the transmission errors remaining at the completion of a manual line-up occur chiefly because of the fixed and manual equalizers. It should be noted however that, when the dynamic regulators are restored to operation after the manual adjustment, any residues will be seized by the dynamics and regulated.

As yet, the field experience with the regulation system and the dynamic shape performance is quite limited. However, the stability of the regulation system and the action of the computer have been well established. It would appear that the major remaining regulation system problem will be the determination of the cause shapes to the accuracy required for long television systems.

Somewhat more experience has been gained with the cosine equalizers. Using a fixed equalizer design based on ten line amplifiers from initial production, a 100-mile circuit, equalized and using 15-cosine terms, yields the residues plotted in Fig. 17. The ripple at the extreme high end of the band is largely caused by the failure of the fixed equalizer to match the very sharp cut-off of the line between 8.35 and 8.50 megacycles. Since telephone channels are not transmitted in this region, and since television requirements are less severe at such high video frequencies 4.2 megacycles, this effect does not limit the transmission quality. Also note that long television circuits, over 400 miles, will have a further level of equalization *D*. There is no visible impairment of ordinary television pictures as a result of insertion of the 200-mile *L3* line in their path. With critical types of test patterns there is a slight effect. While there are problems yet to be solved before 4,000-mile transmission can be obtained, the 200-mile performance is most encouraging.

In the past the adjustment of manual equalizers has often been a difficult and time-consuming task. The cosine adjusting set described previously appears to have made sizable inroads on the solution of this problem. The adjustment of the 25 controls used for television is a 3-minute job. The fact that the equalization operator is working toward a unique solution and therefore knows when he is done appears to be of material value.

Regulation

General

The *L3* regulation system is in many respects similar to the *L1* system. However, the design of a stable regulation system for over 500 regulators in tandem introduces unusual problems. Also the accuracy and stability requirements have led to the use of novel regulator circuits including, for example, an analogue computer as an element of the system.

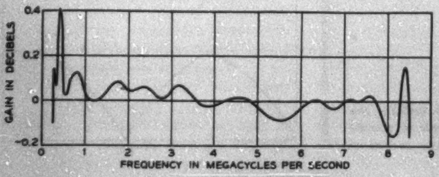


Fig. 17. Final gain characteristic of a 100-mile L3 line after equalization with 15 cosine shapes

Six pilots are used, of 308, 556, 2,064, 3,096, 7,266, and 8,320 kc. These frequencies were selected as the best for measuring the anticipated system changes as restricted by where signal allocations would permit their insertion. For example, the wide gap from 3,096 to 7,266 is largely caused by the difficulty of inserting and removing pilots in the lower video frequencies of television signals. The problem of finding a satisfactory set of pilot levels and frequencies which will at the same time be compatible with the desired signals is an important part of the system design problem.[1]

The change of 4-mile cable loss with temperature is so large (±1.2 decibels) that regulation is required at each repeater. It takes 3 months or more for the cable loss to change 2 decibels but the normal line maintenance interval is of this order. A gain error of this magnitude could not be allowed to accumulate over very many repeaters before the signal-to-noise performance of the system would collapse. Other effects such as vacuum-tube aging and repeater-temperature changes can be allowed to accumulate over as many as 30 repeaters before regulation. These facts dictate the location of regulators in the system. At each repeater there is a line regulator controlling a square-root-of-frequency-shape regulating network. Then at equalizing points and dropping points office regulators correct for the remaining effects.

Chain Action

Pilot-controlled dynamic regulators derive much of their advantage from the fact that they prevent gain changes from accumulating from repeater to repeater. This advantage is one manifestation of what might be called the chain action of a series of regulators. There is however the corresponding disadvantage of disturbances of the pilot causing the accumulation of unwanted gain fluctua-

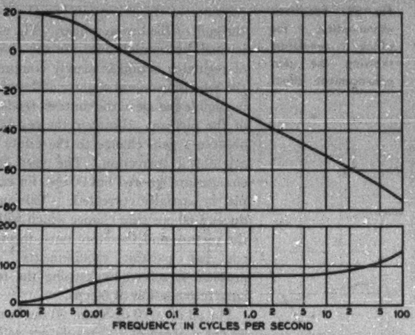


Fig. 19. Loop gain characteristic for the line regulator

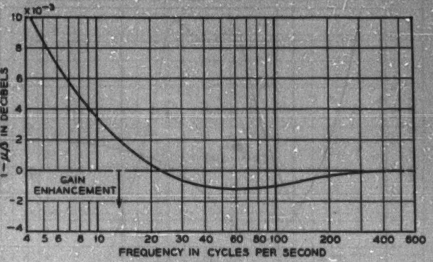


Fig. 20. Feedback characteristic for the line regulator showing the gain enhancement effect

tions. In previous systems this disadvantage has been aggravated by positive envelope feedback ($1-\mu\beta$ less than one) at some frequencies, an effect known as gain enhancement. In the *L3* system the gain enhancement is nearly negligible but the television requirements still require careful control of certain types of gain fluctuations.

The advantage noted can easily be demonstrated by a simple example. Consider a chain of regulators each having 20-decibel envelope feedback so that pilot level changes are reduced by 10 to 1. Now consider what happens if each cable section changes loss by 1 decibel. Table I illustrates the action.

The first regulator inserts a gain change of 0.9 decibel in response to the 1.-decibel input change. The 0.1 decibel-error increases the input change to the second regulator to 1.1 decibel and it therefore inserts a 0.99-decibel correction. The total resultant error of 0.11 decibel adds to the change at the third regulator input, etc. Simply stated, the error of the first regulator rides through the system forcing the other regulators to make an accurate correction. Actually, of course, this statement is oversimplified, but it should be clear that the effective feedback of the regulation system is the (voltage) sum of the feedbacks of the individual regulators. Thus 100 regulators each having 20 decibels of feedback tend to act as a single regulator having 60 decibels of feedback. The rigorous treatment of these effects will be developed later.

Table I. Example of Chain Action

Regulator Number	Input Pilot Change	Inserted Correction	Output Pilot Change
1	1 decibel	0.9 decibel	0.1 decibel
2	1.1	0.99	0.11
3	1.11	0.999	0.111
4	1.111	0.9999	0.1111

Fig. 18 shows a block diagram of a regulator in a form intended to indicate the feedback structure. The feedback loop includes a pilot pickoff filter, amplifier, and rectifier. This converts the output pilot level into a direct voltage. The battery, or actual input signal for the circuit, represents the equivalent of the desired pilot output level. The signal applied to the d-c amplifier is a d-c signal representing the error in pilot level. This d-c signal is, in effect, converted back to a pilot level by the action of the regulating network and its modulation of the input pilot level. Thus changes in input pilot level are equivalent to gain changes in the μ circuit of the feedback structure and are resisted by feedback action just as in any other feedback amplifier. It is also valuable to note the respective μ and β roles played by the various components since the stability requirements, etc., then become clear. For example, the pilot amplifier is in the Beta circuit and therefore must be a highly stable device. On the other hand, the d-c amplifier is in the μ circuit and its drifts are reduced by the loop feedback.

Having developed the feedback nature of the structure and the roles of the components, the conventional feedback art can be used for the analysis of the individual regulator. It can be shown that

$$\frac{\text{Change in output pilot level}}{\text{Change in input pilot level}} = \frac{1}{1-\mu\beta} \qquad (23)$$

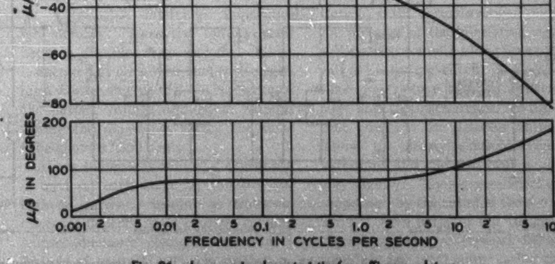


Fig. 21. Loop gain characteristic for office regulators

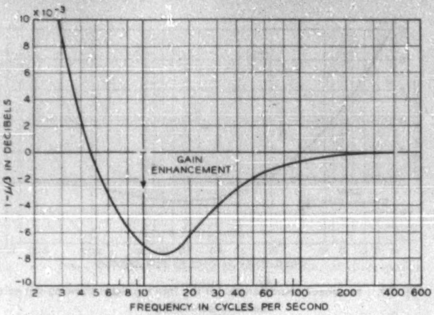


Fig. 22. Feedback characteristic for office regulators showing the gain enhancement effect

$$\frac{\text{System gain change to pilot frequency}}{\text{Change in input pilot level}}=\frac{\mu\beta}{1-\mu\beta} \qquad (24)$$

This result is not very surprising but it can be used to determine the performance of the following regulators in a chain. Many different cases must be considered. Sometimes the pilot levels change because of an effect distributed all along the system. In other cases the change occurs only at the input to the line. Sometimes the pilot level changes are the important effect. In other cases the importance resides in the gain change to the signals. In all cases the results may be complicated by the fact that $\mu\beta$ is, in general, a complex number and thus phase as well as amplitude is important.

Adopting the notation

ΔP_{in} = fractional change in input pilot at nth regulator
ΔP_{on} = fractional change in output pilot at nth regulator
ΔG_n = gain change to signals (near pilot frequency) of nth regulator
ΔG_t = total system gain change $=\Sigma G_n$

one can readily show the following:

Case I. Disturbance of pilot only at input to system:

$$\frac{\Delta P_{on}}{\Delta P_{i1}}=\left[\frac{1}{1-\mu\beta}\right]^n \qquad (25)$$

$$\frac{\Delta G_n}{\Delta P_{i1}}=\mu\beta\left[\frac{1}{1-\mu\beta}\right]^n \qquad (26)$$

$$\frac{\Delta G_T}{\Delta P_{i1}}=\left[\frac{1}{1-\mu\beta}\right]^n-1 \qquad (27)$$

Case 2. Equal gain change in each regulating section:

ΔP_c = fractional gain change of section

$$\frac{\Delta P_{on}}{\Delta P_c}=\frac{1}{\mu\beta}\left[\left(\frac{1}{1-\mu\beta}\right)^n-1\right] \qquad (28)$$

$$\frac{\Delta G_n}{\Delta P_c}=\left[\left(\frac{1}{1-\mu\beta}\right)^n-1\right] \qquad (29)$$

$$\frac{\Delta G_T}{\Delta P_c}=\left\{\frac{1}{\mu\beta}\left[\left(\frac{1}{1-\mu\beta}\right)^n-1\right]-n\right\} \qquad (30)$$

As an example of the application of these equations, consider the effect of television-induced compression. The presence of the television signal reduces the gain of the line amplifier. The effect is small but cumulative. In the absence of regulator action it merely compresses the television signal slightly and makes a negligible change in the contrast rendering of the picture. However the regulators observe a gain change to the pilots and attempt a correction. The very rapid changes are ignored but 60 cps, for example, is partially corrected. This introduces a 60-cps gain change which will lag the picture and therefore must meet 60-cps bar pattern requirements. This problem is solved by keeping the regulator response low at 60 cps.

For $\mu\beta$ of -70 decibels, 90 degrees, at 60 cps a chain of 700 regulators will insert a total gain change approximately one-tenth that of the total compression. If the $\mu\beta$ were allowed to approach -50 decibels, the total gain change would equal the compression, and certain types of pictures would be degraded.

This example brings out one of the important facts: When designing regulators for long systems the $\mu\beta$ characteristic must be carefully controlled to losses much higher than is customary in amplifier design. In a conventional feed-back-amplifier loop-cutoff the magnitude and phase of $\mu\beta$ are no longer of much interest after the magnitude drops below -10 decibels. In *L3* regulators the loop is of vital interest to losses of the order of 70 decibels. This is of course largely due to the fact that the chain action increases the effective system feedback by nearly 60 decibels. Thus the over-all system is similar to conventional amplifier practice. Loop gain $\mu\beta$ and feedback $1-\mu\beta$ characteristics for the line and office regulators are shown in Figs. 19, 20, 21, and 22. Note that 1,000 line regulators in tandem

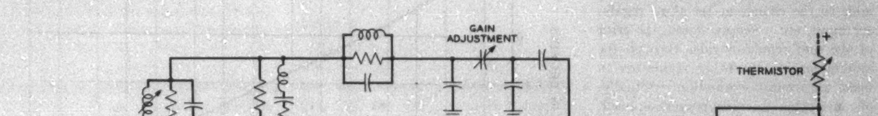


Fig. 23. Simplified schematic of the line dynamic regulator showing the signal circuits

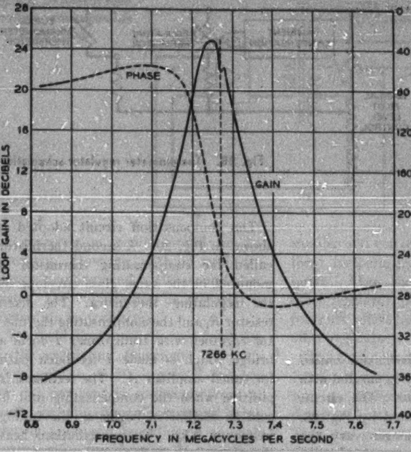


Fig. 24. Loop gain and phase of the 7,266-kc pilot amplifier

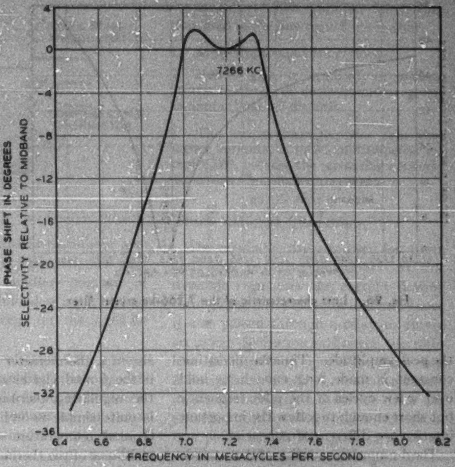


Fig. 25. Regulator selectivity without the crystal filter

give an over-all gain enhancement of only 1.2 decibels. This would be even less it it were not for a 100-cps roll-off in the d-c amplifier to reduce noise. One hundred office regulators give an 0.8-decibel gain enhancement even with their 20-cycles per second roll-off.

THE DYNAMIC LINE REGULATOR

As indicated in Fig. 23, a crystal filter is used to pick the pilot off the line in the presence of the other signals. The filter impedance goes through resonances caused by the crystals, and introduces a gain characteristic on the line that cannot, in practice, be equalized. Thus it is necessary to hide the filter from the line with loss. To hold the transmission distortion of signals to 0.15 decibel with 500 regulators, or 0.0003 decibel per regulator, requires a voltage loss of 23 decibels with the filter impedance changing by large factors from its nominal 25,000-ohm level. The power loss bridging on the 37.5-ohm (75 into 75) circuit is, of course, much greater.

The 7,266-kc pilot level at the line amplifier output is 16 decibels below 1 milliwatt into 75 ohms. The filter and pad loss totals 24 decibels (voltage ratio) leaving an available pilot signal of about 0.002 volt in 25,000 ohms. To solve drift and stability problems in the d-c circuits, the pilot is converted to a d-c voltage of 60 volts. This requires an amplifier-rectifier of a 90-decibel voltage gain, stable with time and temperature.

As indicated in Fig. 23, the amplifier consists of three stages using *403B* (long-life *6AK5*) tubes. Feedback is taken shunt-shunt, output plate to input grid, to stabilize the input and output tuned circuits against Q changes with temperature. The regulators are designed to operate from −20 to +160 degrees Fahrenheit and the amplifier gain does not change by more than 0.3 decibel over this range.

The β circuit contains an adjustable condenser divider for field gain adjustment. The tuned sections are heavily damped to get temperature stability, but nevertheless compensate for the cutoffs of the μ circuit. The resultant loop gain and phase are shown in Fig. 24. It will be noted that the phase margin against singing is rather large, being about 60 degrees. This permits tube replacement without retuning as well as protection against tuning changes caused by time and temperature.

Grid-plate capacity places a limit on the permissible interstage impedance for stability. Thus the output circuit between the third stage and the rectifier is operated as a reactive transformer giving a voltage step-up of 2. This increases the output voltage obtainable without raising the impedance facing the third tube above 12,500 ohms.

About 16 decibels of local d-c feedback is used on each stage to stabilize the cathode current. Insofar as transconductance depends upon cathode current, transconductance changes are reduced. This is of material value in further reducing the effects of vacuum tube aging.

Fig. 25 shows the external gain of the 7,266 kc amplifier without the crystal filter. The filter response is shown in Fig. 26. The relatively wide 3-decibel bandwidth of ±1.5 kc is to reduce the contribution of the filter to the gain enhancement problem. The large rejections to frequencies further removed from the pilot prevents operation of the regulator by signals other than the pilot. Also note that for very strong signals such as the television carrier at 4,139 kc the filter is aided by the amplifier selectivity.

The diode detector is also designed to reduce the effects of interference. The time constant of the detector is made short so that the output will follow envelope fluctuations up to about 40 kc. Thus the output of the detector in the presence of an interfering signal within 40 kc of the pilot becomes the power sum rather than the voltage sum of the two signals. This is readily understood from Fig. 27. Here E_p is the normal rectified pilot. In the presence of the interfering signal E_i a diode detector with a long time constant will deliver a d-c output of E_p+E_i, and thus give voltage addition between the pilot and the interference. With a fast time constant the detector output can follow the nearly sinusoidal envelope variations and the d-c level is changed only slightly. The a-c component may be suppressed by the cutoff of the d-c amplifier but, even if this is not the case, the thermistor being a thermal device responds to the total power rather than to

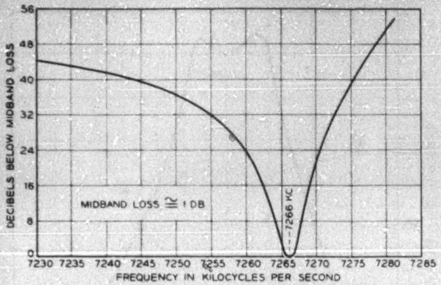


Fig. 26. Loss characteristic of the 7,266-kc crystal filter

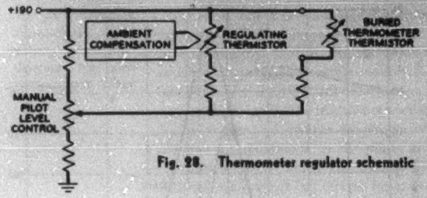


Fig. 28. Thermometer regulator schematic

the peak amplitude. Thus the diode time constant is made long enough to hold over a few cycles of the pilot frequency but short enough to follow the important interference difference-frequencies.

The d-c amplifier consists of three triode sections essentially in parallel but with the biases and feedbacks differing for the purpose of providing a voltage-current characteristic that corrects for sensitivity changes of the thermistor with operating current. This maintains the over-all loop feedback relatively constant over the 1-to-20-milliampere current range. The thermistor is directly heated by the plate current of the d-c amplifier to obtain single time-constant performance of the thermistor. The thermistor transmission, plate current changes as an input and pilot level as an output, is the main frequency characteristic of the regulator loop.

LINE THERMOMETER REGULATORS

It is possible to dilute the regulation system with less costly, less accurate regulators without undue loss of over-all performance. This is accomplished by the use of thermometer regulators at alternate regulating points. These consist of a thermometer thermistor buried in the ground, electrically in parallel with the regulating thermistor. The circuit is quite simple as indicated by Fig. 28. Ground temperature changes vary the resistance of the thermometer thermistor, thereby changing the current and resistance of the regulating thermistor. The manual control is used to effect initial alignment of the system. The regulating sensitivity is designed to overcompensate slightly for cable loss changes to ease the burden somewhat on the following dynamic regulator.

AMBIENT TEMPERATURE COMPENSATION

Both types of line regulators require the assistance of ambient temperature compensation of the regulating thermistor. Conventional compensation circuits would hold the thermistor resistance within about 20 per cent but this would produce an error of 1 decibel at a thermometer regulator, and about 0.2 decibel at a dynamic unit. Thus an improved compensation scheme was required which would connect only to an indirect heater, the bead itself being already controlled by d-c heating from the regulators.

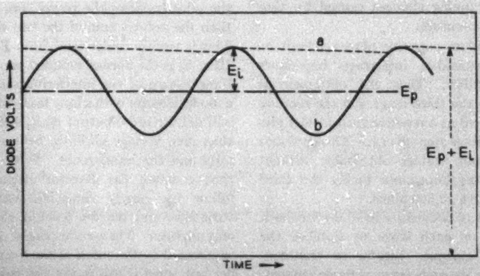


Fig. 27. Diode detector output in presence of interference E_L. E_p is normal pilot signal. Curve a obtains with long time-constant, b with short time-constant

The compensation circuit adopted is shown in Fig. 29. A second thermistor called the compensating thermistor is mounted in the same glass envelope as the regulating thermistor. The fixed resistor R_1 and the compensating thermistor together with transformer T form a bridge which is made a feedback path for tuned amplifier A. The feedback is positive when the compensating unit is cold so oscillation begins at the tuning frequency 4 kc. These oscillations heat the thermistor and tend to bring the bridge into balance. The bridge stabilizes at a small unbalance just sufficiently to yield a loop gain of unity. The level of oscillation is forced to that value which will maintain this small unbalance. Any changes in balance thereafter produce deviations of loop gain from unity and the oscillation level increases or decreases until the equilibrium is re-established. Thus the level of oscillation changes with temperature but the resistance of the compensating thermistor is held constant.

Because the circuit supplies nearly perfect temperature compensation to the unit in the bridge, a suitable fraction of the oscillator power may be fed to the heater of the regulating thermistor to achieve very close compensation of it. Resistors R_2 and R_3 are adjusted in manufacture to correct for slight differences between the two thermistors. Note that the compensating thermistor is provided with an unused heater to match the thermal properties of the two units.

The amplifier consists of a single *403B* tube with 20-decibel d-c feedback for current (and transconductance) stabilization. This feedback is vital because it also introduces a slight compressive action in the amplification of the tube and thereby prevents rapid wild changes in oscillation level. Bypassed d-c feedback on an amplifier causes d-c second-order distortion to increase the bias and thereby reduce the transconductance. This effect overcomes the tendency of the third-order distortion to create expansion in this particular tube. If the thermistor response were fast compared to the reciprocal of the bandwidth of the amplifier,

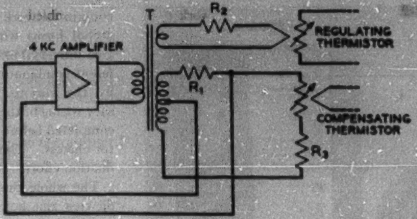


Fig. 29. Ambient temperature compensation oscillator

the compression action would be unnecessary. However, with an audio-frequency amplifier and a 100-second thermistor, the compression is essential in preventing motorboating.

The field limits on the compensation of the regulating thermistor over the range −20 to +160 degrees Fahrenheit are ±3 per cent in resistance from all causes including manufacture and aging. Specific units can be adjusted to yield compensation to a fraction of 1 per cent.

Office Regulators

The *L3* office regulators are similar in design to the line regulator. However the office regulators operate their regulating networks via an analogue computer, and of course a variety of pilot frequencies are employed. Because signals are dropped at offices, higher loop feedbacks are used to insure accurate equalization. However, temperature variations are smaller and conventional thermistor ambient temperature compensation is adequate. Also the smaller number of office regulators permits less isolating loss for nick effect (except for the 7,266-kc office regulator).

The lower levels of the pilots (except 7,266) are compensated by reduced isolation loss, 12 instead of 23 db, and reduced detector level, 40 instead of 60 volts. Thus the gain required is not substantially increased. The pilot amplifier design is therefore different primarily in the tuning frequency and in the simplifications in the lower frequency units permitted by the higher permissible interstage impedances.

The diode detector feeds a cathode follower to obtain the d-c voltage representing the deviation of the pilot from its assigned value as a low-impedance source to feed the computer. The appropriate signals from the computer are fed to the d-c amplifier. This amplifier differs from that used in the line regulators in that:

1. A push-pull input is provided.
2. Higher gain is required to produce greater feedback, 30 decibels, and overcome computer losses, 5 decibels.
3. The output stage supplies somewhat higher currents, 1 to 30 milliamperes, except 7,266, because the regulating networks use a lower impedance thermistor.

Other Regular Functions

The regulators are also used for alarm and pilot indicator functions. At line dynamic regulators the current flowing through the diode detector load resistance is also passed through a relay to obtain an alarm indication whenever the pilot level deviates from normal by more than 3 decibels. At offices, similar arrangements operating on a 2-decibel error are provided both for alarm and switch initiation purposes. If any pilot deviates by 2 decibels the service switches to the spare line. In addition, fast switch initiation is obtained from the 7,266-kc regulator by direct connection to the detector output. This arrangement avoids the time delay of the relay operation. For pilot level indication the diode load current is read on appropriate meters. This avoids the necessity of providing separate pilot level indicators and is possible because of the reliability and stability of the regulator amplifiers.

Fig. 30. Line dynamic regulator as seen from wiring side

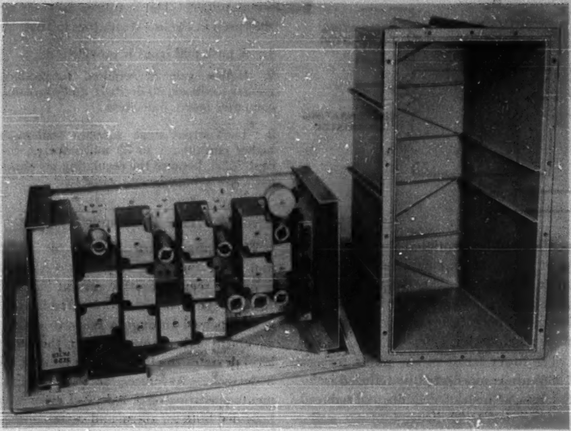


Fig. 31. Top view of line dynamic regulator showing case

MECHANICAL FEATURES

Figs. 30 and 31 are views of a regulator seen from the wiring side and from the top respectively. The chassis consists of two steel end plates riveted to steel angles, with a punched copper plate screwed to this structure. The salient feature of this type of construction is that one universal punched plate can be used for all the regulators, including the thermometer regulator, any individual regulator being fabricated by mounting the necessary component cans and vacuum tube sockets on it. Power wiring and all leads that are not critical as to length or placement can be run in the wiring trough around the edge of the chassis shown in Fig. 30. This eliminates the necessity of lacing the wires into a cable, a significant saving in production effort.

The component cases are shown in Fig. 31. They are zinc die castings and contain network elements assembled on stypol forms which fit inside the cans. One universal case accommodates 64 different combinations of elements required by the various regulators. The necessary wiring of the individual cases may be completed before assembly on the regulator chassis. This feature also saves production effort.

The whole chassis is mounted inside a die-cast zinc housing, and all power and test leads are brought into the regulator through airtight connections. The two parts of the housing when assembled together are made airtight by a rubber gasket which fits into the slot around their inner edges. The general construction features and size of the regulator assembly can best be understood by inspection of the photographs.

References

1. THE *L3* COAXIAL SYSTEM, **C. H. Elmendorf, R. D. Ehrbar, R. H. Klie, A. J. Grossman.** *AIEE Transactions*, vol. 72, pt. I, Sept. 1953, pp. 395-413.

2. *L3* COAXIAL SYSTEM—AMPLIFIERS, **L. H. Morris, G. H. Lovell, F. R. Dickinson.** *AIEE Transactions*, vol. 72, pt. I, (*Paper 53-222*).

3. VARIABLE EQUALIZER, **H. W. Bode.** *Bell System Technical Journal*, April 1938, vol. 17, pp. 229–44.

4. ATTENUATION AND DELAY EQUALIZERS FOR COAXIAL LINES, **W. R. Lundry.** *AIEE Transactions*, vol. 68, pt. II, 1949, pp. 1174–79.

5. VARIABLE ATTENUATION NETWORK, **P. H. Richardson.** United States Patent No. *2,096,028*.

6. INTRODUCTION TO THE THEORY OF FOURIER SERIES AND INTEGRALS (book), **H. S. Carslow.** The Macmillan Company, New York, N. Y., third edition, 1930, p. 148.

No Discussion

Electron-Ion Recombination at Low Pressures

SIDNEY BOROWITZ
NONMEMBER AIEE

AS THE name implies, the recombination process describes the situation in which a neutral gas is ionized and the electrons and ions which are thus formed recombine to form neutral atoms and molecules again. Its importance in the field of gas discharge derives from the fact that this process constitutes one of the important agencies whereby charges disappear from a discharge. From the point of view of physics in general, it is one of the fundamental processes which all the artillery of classical and modern physics must be capable of explaining.

The pioneering work in this field has been done by Kenty[1] and Mohler,[2,3] and has been summarized by Loeb.[4] Since these experiments were performed, and especially in recent years, a considerable amount of literature has appeared relating to the recombination process. It is not the purpose of this paper to summarize all of this work, but rather to discuss those experiments which seem to shed most light on the nature of the fundamental process involved.

The recombination process is described mathematically by noting that the rate of disappearance of electrons per unit volume must be proportional to the product of the density of electrons and the density of positive ions. If, as is almost always the case, the density of electrons and positive ions are equal, we have

$$\frac{dn}{dt} = -\alpha n^2 \tag{1}$$

Paper 53-263, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 3, 1952; made available for printing April 20, 1953.

SIDNEY BOROWITZ is with New York University, New York, N. Y.

The research reported in this paper has been sponsored by the Geophysics Research Division of the Air Force Cambridge Research Center, Air Research and Development Command, under contract no. *AF-19(122)-463*.

where

n = the number of electrons per cubic centimeter
α = the recombination coefficient, which measures the rate at which the process proceeds

If α is not a function of n, equation 1 can be integrated to give

$$\left(\frac{1}{n}-\frac{1}{n_0}\right)=\alpha t \qquad (2)$$

where n_0 is the electron density at time $t=0$. It is seen, therefore, that the problem of the measurement of the recombination coefficient resolves itself to a problem of the measurement of n as a function of t. Basically, all of these experiments are performed in identical ways. Some gas in a discharge tube is ionized and the electron density is measured in the tube as a function of the time after the power has been turned off. Of course, in the interpretation of the experiment, one must be certain that the electrons are really lost by recombination. This means that some account must be taken of the electron loss due to processes which compete with recombination. For example, n may change, due to diffusion to the walls or attachment to neutral atoms, molecules, or impurities. The experiments which are most easily interpreted are those done with gases which do not attach electrons very readily and are of high purity; also, the experiments should be either at pressures high enough so that wall diffusion does not contribute significantly to electron loss, or an evaluation of the contribution of wall diffusion should be possible.

If one considers a collision of a positive ion and an electron, it is clear that recombination cannot take place if there are no other forces acting on this system. For if one examines the collision in the coordinate system in which the center of gravity is at rest before recombination, one sees that the center of gravity must remain at rest after the collision; that is, in order to conserve momentum, the neutral atom or molecule must be at rest in this system. However, if this were true, since the forces between these particles are attractive, and the potential energy of the combined system is less than that of the particles when they are separated, we could not satisfy both the conservation of energy and momentum laws. Therefore, in order for recombination to take place, a third body must be present to absorb some additional energy, if the two basic conservation laws are to be satisfied simultaneously. This third body can be only some complex of the neutral gas, the container walls, another charged particle, or a light quantum. A knowledge of the properties of the recombination coefficient allows us to discriminate between these possibilities.

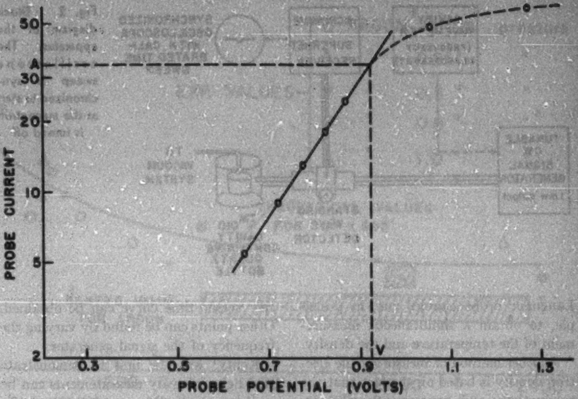


Fig. 1. Probe characteristics

For example, if α is pressure independent, the third body cannot be a neutral constituent of the discharge. If one finds experimentally that equation 2 is satisfied, we can be reasonably certain of two things:

1. That diffusion to the walls is an insignificant process;
2. That the third body is not another charged particle.

For if the former were the case, dn/dt would be proportional to n, and if the latter were true, it would be proportional to n^4; in neither case would the reciprocal of the electron density be proportional to t. If one could decide that a photon must be the third body, either by the elimination of any other possibility, or by demonstrating in some other way that every recombination leads to the emission of light, one has still another way of measuring the recombination coefficient, namely by measuring the light intensity as a function of the time after the discharge has been turned off.

With this introduction, we turn now to a discussion of the various measurements of α as a function of temperature and pressure for various gases.

Basic Experimental Techniques

As indicated in the preceding section, the measurement of the recombination coefficient involves basically a measurement of the electron density in a decaying plasma as a function of the time, or possibly it involves a light-intensity measurement. In this section, we shall discuss briefly the various possibilities for these measurements.

One of the oldest methods of electron charge-density measurement still in use today is an outgrowth of an idea of Langmuir.[5,6] If one inserts a small probe into the discharge, the current to the probe i_p is given by the formula

$$i_p=A\epsilon^{\frac{v-V}{kT}} \qquad (3)$$

where

$$A=Nea\frac{ekT}{2\pi m}$$

where

N = electron density
a = area of the probe
v = potential of probe
V = potential of plasma
e,m = charge and mass of electron
T = gas temperature
k = Boltzman's constant

Equation 3 is valid when $0<v-V<<1$ and the velocity distribution of the charged particles is Maxwellian. If one plots the logarithm of the probe current, as a function of the probe potential, the resulting curve is a straight line as long as our hypothesis holds. When $v=V$, the curve will cease to be linear, and the plot of log i_p versus V will show a change in curvature, as shown in Fig. 1. The slope of the straight-line portion will be a measure of the gas temperature, and the value of i_p at which the curve becomes nonlinear, gives the value for A. Thus a determination of the characteristics of the

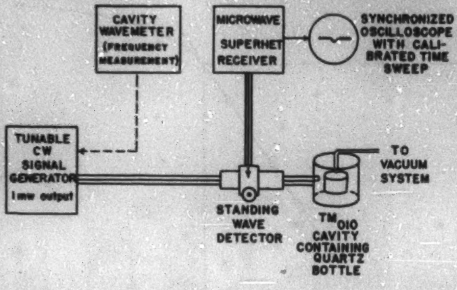


Fig. 2. Block diagram of the apparatus. The oscilloscope sweep is synchronized to start as the magnetron is turned off

Langmuir probe enables one, in principle, to obtain a simultaneous measurement of the temperature and the density.

Another method of measuring the electron density is based on the fact that the conductivity and, consequently, the resonant frequency of a cavity depends on the electron density of the material in it. In fact, Slater[7] has shown that under certain conditions, which are always met in the experiments

$$\Delta\lambda = C\lambda^2 n_a \tag{4}$$

where

λ = wave length corresponding to the resonant frequency of the cavity when it is empty
$\Delta\lambda$ = the change in wave length due to the presence of electrons
n_a = average electron density in the cavity

The coefficient C can be calculated and depends on the geometrical configuration of the system and on the distribution of charges and electric field within the cavity. A typical experimental arrangement using this method is illustrated in Fig. 2.[8]

A tunable continuous-wave signal generator produces a small signal (less than 1 millowatt) which is fed through a slotted line into the cavity containing the bottle which has the decaying plasma. The frequency of the generator differs from the natural frequency of the cavity by a few megacycles. A point on the line can be determined[8] such that a probe placed there will have a minimum amount of power flowing into it when the resonant frequency of the cavity is equal to the frequency of the signal generator. The oscilloscope is triggered to go on when the discharge goes off. By noting at what time one gets a minimum in the oscilloscope trace one can determine at what time the resonant frequency of the cavity was the same as the generator frequency. Thus, since the change of frequency (or wave length) is proportional to the electron density with a known proportionality constant, one point on the electron density versus time curve can be obtained. Other points can be found by varying the frequency of the signal generator.

Kenty[1] was the first to demonstrate that light intensity measurements can be used to measure the recombination coefficient. He noted that for several milliseconds after the power was turned off, light was emitted from the discharge tube. In his experiment with argon he made a spectroscopic analysis of this afterglow, and found lines of the atomic argon spectrum originating from a level 0.084 volt below the ionization level. This is 15.6 volts above the ground state and 3.9 volts above the metastable level. Yet, when a small amount of sodium was added, the D-lines of sodium were not excited. He concluded therefore, that since but 2 volts are required to excite the sodium D-lines, the spectrum of argon could not be due to the acceleration of the electrons by space charge, but is associated with the recombination process.

The methods that have been used to measure the light intensity of the afterglow are standard. Mohler[2,3] used a photometer and a standard source, Holt[10] and Biondi and Holstein[11] a photomultiplier tube with an amplifier.

Results for the Recombination Coefficient

The three most important measurements of α involving the use of the Langmuir probe have been made by Kenty,[1] Mohler,[2,3] and Sayers.[12] Kenty's measurements were made in argon at a pressure of 0.8 millimeters of mercury and at a temperature of 3100 degrees Kelvin. Those of Mohler were made in mercury at a pressure of 0.27 millimeters of mercury and at temperatures ranging from 1560 to 2300 degrees Kelvin, and in caesium at a pressure of 0.11 millimeters of mercury and at temperatures of 1200 to 1360 degrees Kelvin. The values obtained by them varied from 2×10^{-10} to 4×10^{-10} cubic centimeters per second. In both these sets of measurements, the correction for diffusion was determined by measurements of the density of the current to the walls and by then computing the recombination coefficient from the equation

$$\frac{4}{3}\pi r^3 \frac{\Delta n}{\Delta t} = \frac{4\pi r^2}{e} i_w + \frac{4}{3}\pi r^3 \alpha n^2 \tag{5}$$

where

n = particle density
i_w = current to the wall
r = radius of the spherical container used
e = electronic charge.

In addition, Mohler[2,3] attempted to measure the light intensity from the afterglow.

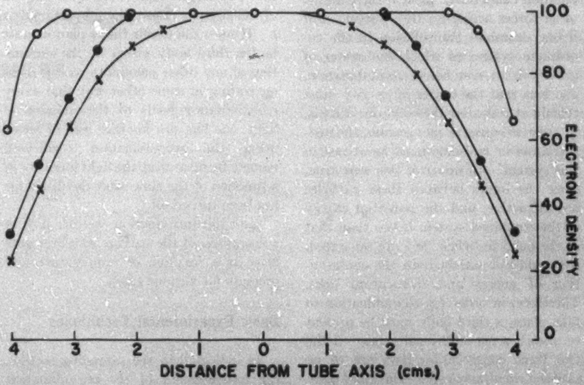


Fig. 3. Distribution of electron concentration across the tube diameter at various intervals of time in the recombination period

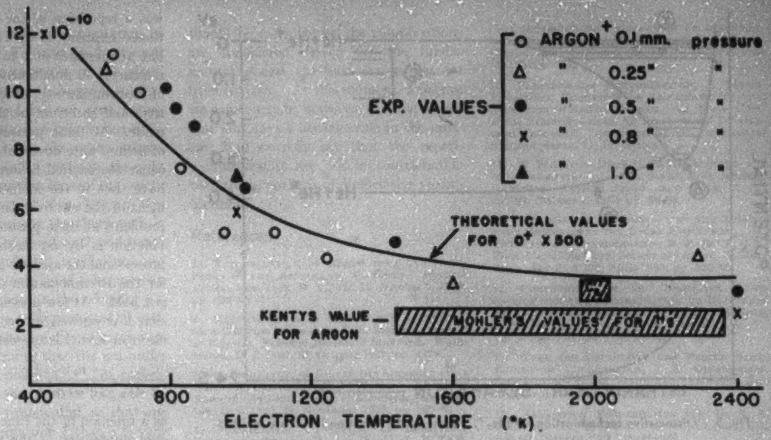


Fig. 4. Combined results of recombination coefficient for various electrons temperatures and gas pressures, and comparison with curve for radiative recombination

If one assumes that the light is due solely to recombination, then the light intensity j would be proportional to dn/dt and consequently to n^2. Thus $n = Kj^{1/2}$. Substituting in equation 2, we have

$$\left(\frac{1}{j_1^{1/2}} - \frac{1}{j_0^{1/2}}\right) = K\alpha t \qquad (6)$$

Mohler found that a plot of $1/j^{1/2}$ versus time was indeed a straight line for several milliseconds, and the slope of this line determines $K\alpha$, according to equation 6. K is then independently determined by comparing the electron density as measured by a Langmuir probe and a light intensity measurement at some particular time. The values of α thus obtained agree with the probe measurements.

Sayers[12] has made measurements in argon at pressures ranging from 0.1 to 1 millimeter of mercury, and at temperatures ranging from 600 to 2400 degrees Kelvin. He has employed a novel way of correcting for wall diffusion. Instead of using a single probe, he uses approximately twelve of them spaced along a diameter. A record for a typical run is summarized in Fig. 3.[12] He attributes the more rapid decrease of electron density near the walls as compared to its decrease at the center, to wall diffusion. If one restricts oneself to times such that the "plateau" has not "eroded" to the center of the discharge tube, no correction for wall diffusion is necessary. Only the decrease in electron density at the center of the discharge tube is used to determine the recombination coefficient. Sayers found no pressure dependence over the range of pressures examined. His results, as well as those of Kenty[1] and Mohler,[2,3] are summarized in Fig. 4.[12]

The experiments measuring the recombination coefficient by using the resonant-frequency measurement as described in "Basic Experimental Techniques" have been carried out by Biondi and Brown[9] and Holt[10] and his coworkers. Since there is no possibility of an independent temperature measurement using this method, it is necessary to wait until the electrons have cooled down to the temperature of the gas. This does not allow as wide a variation of temperature measurements as does the probe technique. Some experiments have been done, however, with the gas cooled to as little as 77 degrees Kelvin and as high as 410 degrees Kelvin. The results of these investigations are consistent with each other. Although a wide variety of gases have been used, we shall mention only the results for the noble gases.

Biondi and Brown[9] have measured the electron density as a function of the time, using the resonant-cavity technique, after ionizing the gas by an electrodeless microwave discharge. By plotting the reciprocal density as a function of the time, one can deduce from the slope of the curve, if it is a straight line, the value for α according to equation 2. To insure a straight line for the plot, it is necessary to work at relatively high pressures. Their results at 300 degrees Kelvin over the pressure range 15 to 30 millimeters of mercury are summarized in Table I.

Holt[10] has used the same technique, but has added a light measurement by means of a photomultiplier tube. He calibrates the photomultiplier by a density measurement at some particular time, and finds that the recombination coefficient for neon is substantially the same as those reported above.

The measurements using the Langmuir probe and those using the resonant cavity differ from each other by several orders of magnitude, even though each group is consistent within itself. It is true, of course, that the same temperature range has not been covered by each and that the measurement of the recombination coefficient by the Langmuir probe method at 300 degrees Kelvin might remove the discrepancy. It is not inconceivable that the recombination coefficient is an extremely sensitive function of temperature between 300 to 600 degrees Kelvin. However, in evaluating which of the results is more nearly correct it might be well to keep in mind the fact that the resonant cavity method is simpler, from a theoretical point of view, than the Langmuir-probe method, and the latter method involves additional difficulties in measurements of this sort, as pointed out by Johnson and Malter.[11]

Comparison with Theory

The indication in the experiments that α is independent of pressure suggests very strongly that the mechanism is essentially

Table I. Reciprocal Densities[9]

Gas	Cubic Centimeters per Ion Second
He	1.7×10^{-8}
Ne	2.08×10^{-7}
A	3×10^{-7}

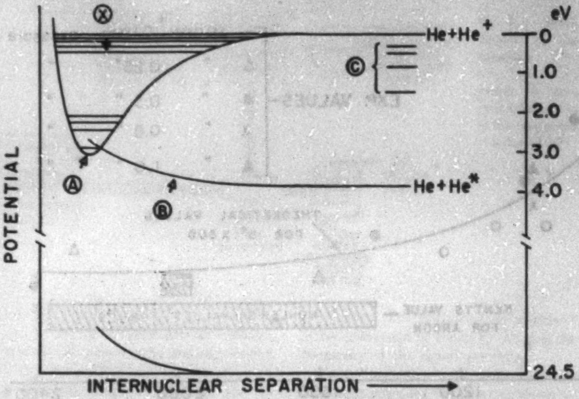


Fig. 5. Dissociative recombination for He_2^+. Some of the vibration levels are omitted

a two body process accompanied by the emission of a photon. Theoretical estimates of this type of recombination process have been made by Bates[14] for oxygen atoms, by Stueckelberg and Morse[15] and Oppenheimer[16] for hydrogen atoms, and by Goldberg[17] and Huang[18] for helium. If one uses the probe measurements, the results differ from the theoretical estimates by a factor of 500. For the others, the discrepancy lies between 10^3 and 10^5. While the theory for the most part deals with gases not used in the experiments, and some of the theoretical estimates are necessarily rough, the magnitude of the discrepancies suggest strongly that the mechanism of the recombination is not simply radiative in the experiments that have been done. The care that the experimenters have exercised in using pure gases, and the fact that the recombination coefficient is independent of the pressure, poses a real dilemma in understanding by what process as fundamental a phenomenon as electron-ion recombination takes place.

Recently, Bates[21] has suggested that the molecular ion X_2^+ plays a fundamental role in the recombination process, and that the mechanism is $X_2^+ + e \rightarrow X + X'$, which he called dissociative recombination. Even though the molecular ion must be formed by a three-body collision, the recombination coefficient would be pressure independent, provided that the atoms recombine much more slowly than the molecular ions are formed. The problem of calculating α for this process quantum mechanically represents enormous difficulties, and the best hope of discussing the reasonableness of the hypothesis is further experimentation.

Further Experiments and Conclusions

In an effort to learn more about the recombination process the microwave-breakdown technique was combined with an analysis of the recombination spectrum.[19] In the case of helium at pressures of 10 to 27 millimeters of mercury, it was found the most prominent spectral lines consisted of He_2 bands. This phenomenon is consistent with dissociative recombination and subsequent formation of the excited helium molecule by a three-body collision. It may also be due to radiative recombination and subsequent formation of the excited helium molecules. The latter possibility, however, has difficulties discussed in the section entitled "Comparison with Theory."

One could test the dissociative recombination hypothesis further by reducing the number of processes that could possibly take place in a discharge. For example, if the pressure of helium were so low that no subsequent formation of He_2^+ were possible, one could learn more about the products of the dissociation, since analysis of atomic spectra is simpler than molecular spectra. By reducing the pressure still further, it might be possible to inhibit the formation of molecular ions altogether, and thus the role played by dissociative recombination could be analyzed.

Biondi and Holstein[11] and Biondi[20] have recently carried out two experiments with these objectives. The first of these was a repetition of the previously mentioned experiment by Holt with helium but at pressures of 1 to 5 millimeters of mercury. A simultaneous determination of electron density, absolute light intensity, and spectrum of the afterglow was made. At these pressures, atomic lines of helium were observed, presumably because the excited helium atoms did not have time to recombine. The measurement of the electron density poses some problems at these pressures since recombination is by no means the dominant process and the analysis given in "Results for the Recombination Coefficient" does not hold. At low pressures, where diffusion loss outweights recombination, the electron density is approximately given by

$$\frac{n}{1+\alpha T_n} = \frac{n_0}{1+\alpha T n_0} e^{-\frac{t}{T}} \tag{7}$$

as a function of the time. n_0 here is the electron density at the time $t=0$. T is the mean decay time due to diffusion, and is related to the ambipolar diffusion coefficient D_a and the diffusion length Λ by

$$T = \frac{\Lambda^2}{D_a} \tag{8}$$

Since D_a, and hence T, is known from previous measurements, a measurement of n enables us to determine α. The results of the Biondi and Holstein[11] experiment can be summarized as follows: The absolute light intensity in the afterglow is proportional to n^2. Absolute-intensity measurements indicate that, for each electron lost by recombination, one light quantum is emitted. The radiation in the visible region are atomic lines arising from states, all of which are within 1.5 electron volts of the atomic ionization potential. By referring to Fig. 5[11] one can see what the implications of these conclusions are with respect to the hypothesis of dissociative recombination. The molecular ion cannot be in its ground state, since the repulsive branch of the He, He* potential curve lies below the discrete levels from which the radiation originates. The only possibility of salvaging the dissociative recombination hypothesis is to assume that the molecular ion has energy greater than the discrete levels (C in Fig. 5) and therefore must be in some excited state (such as X in Fig. 5). While this is an extreme hypothesis, it has been estimated by Massey[22] that these high vibration states are metastable, and may persist long enough to participate in the recombination process.

The more recent work of Biondi[20] had as its objective the examination of the recombination process when the pressure was so low as to affect the formation of molecular ions. To this end, Biondi measured the electron loss in the afterglow of a microwave discharge in a mixture of helium and 0.1 per cent argon. The metastable state of helium has sufficient energy to ionize the argon atoms, so that if one waits long enough, the only ions that are present are those of the argon. The measurement of the electron loss enables one to measure the characteristic time constant, T of equation 7, and hence the ambipolar diffusion coefficient; see equation 8. A knowledge of this quantity allows one to compute the mobility μ_+. This value agrees with previously[23] estimated values of the mobility of argon ions in helium, confirming that the ion under consideration is really A^+. The experimental data on the electron loss can be fitted by a pure exponential, so that one concludes from equation 7 that α is very small. A detailed estimate of the experimental errors shows that $\alpha<10^{-3}$ times the recombination coefficient of pure argon. This may mean that in this experiment we have radiative recombination to the atomic ion and its value is 0.001 of the value ordinarily measured, or that we still have recombination to the molecular ion whose concentration is smaller than in ordinary discharge.

We have confined ourselves in this report to those experiments which shed some light on the fundamental mechanism of recombination of ions and electrons. We have omitted many interesting details about peculiar temperature and pressure dependence, since we have felt that the mechanism of the process should be clear before the complications are considered. Even for the fundamental process, the final chapter has not been written. At this moment, I believe we can conclude, however, that the molecular ion plays a decisive role in the process; and we may say that the experimental results are not in contradiction with the hypothesis of dissociative recombination.

References

1. Recombination of Argon Ions and Electrons in the Afterglow, C. Kenty. *Physical Review*, New York, N. Y., vol. 32, October 1, 1928, pp. 624–35.

2. Recombination of Ions in the Afterglow of a Caesium Discharge, F. L. Mohler. *Journal of Research*, National Bureau of Standards, Washington, D. C., vol. 19, October 1937, pp. 447–56.

3. Recombination of Ions in the Afterglow of a Mercury Discharge, F. L. Mohler. *Journal of Research*, National Bureau of Standards, Washington, D. C., vol. 19, November 1937, pp. 559–66.

4. Fundamental Properties of Electrical Discharge Through Gases (book), L. B. Loeb. John Wiley and Sons, New York, N. Y., 1939.

5. Positive Ion Currents in the Positive Column of the Mercury Arc, I. Langmuir. *General Electric Review*, Schenectady, N. Y., vol. 23, November 1923, pp. 731–35.

6. Studies of Electric Discharges in Gases at Low Pressures, I. Langmuir, H. Mott-Smith, Jr. *General Electric Review*, Schenectady, N. Y., vol. 24, July 1924, pp. 449–55.

7. Microwave Electronics, J. C. Slater. *Reviews of Modern Physics*, New York, N. Y., vol. 18, October 1946, pp. 441–512.

8. Measurements of Ambipolar Diffusion in Helium, M. A. Biondi, S. C. Brown. *Physical Review*, New York, N. Y., vol. 75, June 1, 1949, pp. 1700–05.

9. Measurement of Electron-Ion Recombination, M. A. Biondi, S. C. Brown. *Physical Review*, New York, N. Y., vol. 76, December 1, 1949, pp. 1697–1700.

10. Recombination Spectrum and Electron Density Measurements in Neon Afterglows, R. B. Holt, *et al.* *Physical Review*, New York, N.Y., vol. 77, January 15, 1950, pp. 239–41.

11. Concerning the Mechanism of Electron-Ion Recombination, M. A. Biondi, T. Holstein. *Physical Review*, New York, N. Y., vol. 82, June 15, 1951, p. 962.

12. Electron-Ion Recombination, J. Sayers. *Proceedings*, Conference on Ionospheric Physics, Geophysics Research Division, Air Force Cambridge Research Center, Air Research and Development Command, Cambridge, Mass. Geophysical Research papers 11, 12, section K.

13. A Floating Double Probe Method for Measurements in Gas Discharge, E. O. Johnson, L. Malter. *Physical Review*, New York, N. Y., vol. 80, October 1, 1950, pp. 58–68.

14. Dissociation, Recombination and Attachment Processes in the Upper Atmosphere. II. The Rate of Recombination, D. R. Bates, *et al.* *Proceedings*, The Royal Society of London, London, England, vol. A 170, April 1939, pp. 322–40.

15. Computation of the Effective Cross Section for the Recombination of Electrons with Hydrogen Ions, E. C. G. Stueckelberg, P. M. Morse. *Physical Review*, New York, N. Y., vol. 36, July 1, 1930, pp. 16–23.

16. Über die Strahlung der Freien Elektronen im Coulombfeld, J. R. Oppenheimer. *Zeitschrift für Physik*, Berlin-Charlottenburg, Germany, vol. 55, July 1929, pp. 725–37.

17. Transition Probabilities for He I, L. Goldberg. *Astrophysical Journal*, Chicago, Ill., vol. 90, October 1939, pp. 414–28.

18. The Continuous Absorption Coefficient of the Helium Atom, S. S. Huang. *Astrophysical Journal*, Chicago, Ill., vol. 108, November 1948, pp. 354–72.

19. Electron Removal in Helium Afterglows, R. A. Johnson, B. T. McClure, R. B. Holt. *Physical Review*, New York, N. Y., vol. 80, November 1, 1953, pp. 376–79.

20. Concerning the Mechanism of Electron-Ion Recombination, II, M. A. Biondi. *Physical Review*, New York, N. Y., vol. 83, September 1, 1951, p. 1078.

21. Electron-Ion Recombination in Helium, D. R. Bates. *Physical Review*, New York, N. Y., vol. 77, March 1, 1950, p. 718.

22. Collisions Between Atoms and Molecules at Ordinary Temperatures, H. S. W. Massey. *Reports of the Physical Society on Progress in Physics*, London, England, 1948–9, pp. 248–269.

23. The Mobility of Alkali Ions in Gases. I. The Attachment of Water Molecules to Alkali Ions in Gases, R. J. Munson, A. M. Tyndall. *Proceedings*, The Royal Society of London, London, England, vol. A 172, July 1939, pp. 28–42

No Discussion

Equivalent Machine Constants for Rectifiers

I. K. DORTORT
MEMBER AIEE

IN CALCULATING the short-circuit currents on a d-c system and the contribution by a particular machine, the task would be greatly simplified if each connected machine could be replaced by an equivalent machine having a constant electromotive force (emf) and linear internal resistance and inductance. Even the highly ramified 250-volt d-c system of a present-day steel mill would then lend itself to rather accurate analysis.

The characteristics of practical machines are not linear over the entire range from no-load to short circuit, and cannot be accurately matched with equivalent constants, except by sections. This is particularly true of rectifiers, which pass through various modes of operation between no-load and short circuit. Matching equivalent constants to sections of the characteristic is not too great a hardship when solving for steady-state fault currents, but does not lend itself to transient condition studies.

The Report on Protection of Electronic Power Converters[1] treats the calculation of fault currents in rectifiers at short circuit only. The subject of equivalent machine constants was purposely avoided because of the lack of groundwork and

---

Paper 53-271, recommended by the AIEE Electronic Power Converters Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 25, 1953; made available for printing April 17, 1953.

I. K. Dortort is with the I-T-E Circuit Breaker Company, Philadelphia, Pa.

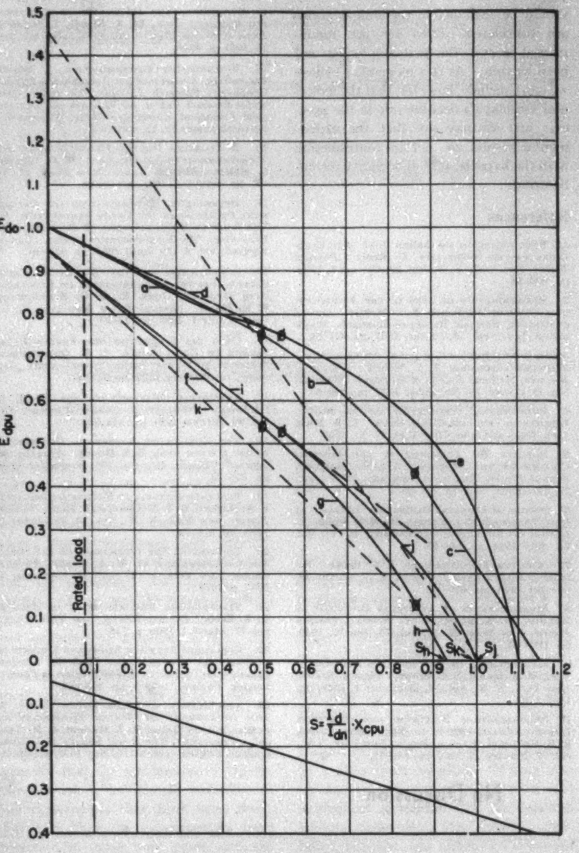


Fig. 1. Regulation curves

Inductive load, no losses: a—1st mode, b—2nd mode, c—3rd mode
Noninductive load, no losses: d—1st mode, e—2nd mode
Inductive load, with losses: f—1st mode, g—2nd mode, h—3rd mode
Noninductive load, with losses: i—1st mode, j—2nd mode
k—Straight-line approximation for fgh
r—Resistance drop line

supporting literature. The voltage regulation curve of rectifiers having only primary reactance had not been worked out except over the normal operating range.

Scope

The first requirement is, then, the complete voltage-current characteristics of rectifiers having little or no secondary reactance, the usual condition met in practice. The necessary equations have been derived and are shown graphically in Fig. 1. These equations and the constants based on them apply only to the 6-phase double-way and double-wye rectifier connections, circuits 23, 45 (and variations) of the American Standards Association standard *C34.1–1949*.[2] The double-way circuit is commonly used in mechanical rectifiers; most low-voltage mercury-arc rectifiers use the double-wye.

No attempt has been made to arrive at a complete system of equivalent constants. It became apparent early in the study that any simplified method having general applicability would lead only to first approximations. Accurate solutions for transient conditions will require laborious step-by-step processes. The treatment is confined to external faults.

Nomenclature

As far as possible the symbols have been made consistent with the nomenclature of the AIEE Committee Report.[1]

E_a = arc drop
E_{ao} = arc drop at no load (projected)
E_d = d-c output voltage
E_{do} = induced direct voltage
E_r = direct voltage drop caused by losses
$\mathcal{E}_o$ = emf of equivalent machine for first mode
$\mathcal{E}_s$ = emf of equivalent machine at or near short circuit
I_d = d-c load current
I_{dn} = rated d-c load current
i_d' = instantaneous transient direct current
K = transient factor $= 1+e^{-3.22}R_c/X_c$
$\mathfrak{L}$ = inductance of equivalent machine
m = 2 for double-way rectifier
= 1/2 for double-wye rectifier
P_r = copper losses at rated load
R_a = equivalent resistance of arc drop
R_c = commutating resistance
$\mathfrak{R}$ = resistance of equivalent machine
$\mathfrak{R}'$ = equivalent transient resistance
$S = X_{cpu}I_d/I_{dn}$, load ratio $=\sqrt{2/3}F_x$ (F_x is the commutating reactance factor defined in the American Standards Association standard *C34.1—1949*.[2] The relation between S and F_x holds for both connections discussed in this paper)
u = angle of overlap
X_c = commutating reactance
Subscript pu denotes per-unit quantities
Subscripts f, h, i, j, and k refer to particular curves

Voltage Regulation Curves

Since both single-way and double-wye rectifiers are being considered here, the terms primary and secondary reactance should be clarified: The current through a primary reactance has no d-c component; the current through a secondary reactance is unidirectional. More definitively, this applies to the leakage fluxes. In most industrial rectifiers the secondary reactance is only a small part of the total and can be neglected.

The direct voltage drop produced in the output of a rectifier by the commutating current acting on the commutating reactances is linear with load current over the normal range when supplying a highly inductive load, and almost linear for a noninductive load. This holds true until the angle of overlap u becomes equal to 60 degrees. Then another mode of operation occurs in which u remains constant. Additional voltage drop is produced by progressive delay in the pickup of the incoming phase in a manner resembling phase control. When supplying a highly inductive load, a third mode occurs in which u again increases to 120 degrees at short circuit.

There is a marked difference in the regulation and sustained short-circuit currents of rectifiers with no inductance in the d-c circuit and those with highly inductive d-c circuits. The d-c circuit can be considered highly inductive if $L_d > 20X_c/2\pi f$ for the double-way rectifier or $5X_c/2\pi f$ for the double-wye rectifier.

In a loss-free rectifier with inductive d-c circuit, from $X_{cpu}I_d/I_{dn}=0$ to 1/2 ($u=60$ degrees)

$$\frac{E_d}{E_{do}}=1-\frac{1}{2}\frac{I_d}{I_{dn}}X_{cpu}=1-\frac{1}{2}S \qquad \text{(curve } a\text{, Fig. 1)} \quad (1)$$

For $S=1/2$ to $\sqrt{3}/2$

$$\frac{E_d}{E_{do}}=\frac{1}{2}\sqrt{3}\sqrt{1-S^2} \qquad \text{(curve } b\text{)} \quad (2)$$

From $S-\sqrt{3}/2$ to $2/\sqrt{3}$ at short circuit

$$\frac{E_d}{E_{do}}=\sqrt{3}-\frac{3}{2}S \qquad \text{(curve } c\text{)} \quad (3)$$

For the case with a noninductive d-c circuit, from $S=0$ to $\sqrt{3}/\pi$ ($u=60$ degrees), E_d/E_{do} is given by curve d of Fig. 1. As a straight-line approximation we can use

$$\frac{E_d}{E_{do}}=1-\frac{\pi}{4\sqrt{3}}S \qquad \text{(curve } d\text{)} \quad (4)$$

From $S=\sqrt{3}/\pi$ to $2\sqrt{3}/\pi$ (at short circuit)

$$\frac{E_d}{E_{do}}=\frac{1}{2}\sqrt{3}\sqrt{1-\left(\frac{\pi}{2\sqrt{3}}S\right)^2} \qquad \text{(curve } e\text{)} \quad (5)$$

The arc drop can be expressed with fair accuracy as the sum of a fixed voltage, which is the arc drop projected to no-load, and a resistance drop which increases linearly with load.

$$E_a=E_{ao}+mR_aI_d \quad (6)$$

where m is 2 for the double-way, and 1/2 for the double-wye rectifier. In most cases E_{ao} can be neglected at heavy overloads.

Copper losses also produce a voltage drop which varies linearly with load, and the total drop caused by losses is

$$E_r=mR_aI_d+\frac{P_r}{I_{dn}^2}I_d+E_{ao} \qquad \text{(curve } f\text{)} \quad (7)$$

Expressing the net voltage-current characteristic in per-unit quantities based on E_{do} and I_{dn}/X_{cpu}, we have

$$E_{dpu}=f(S)-\frac{E_r}{E_{do}} \quad (8)$$

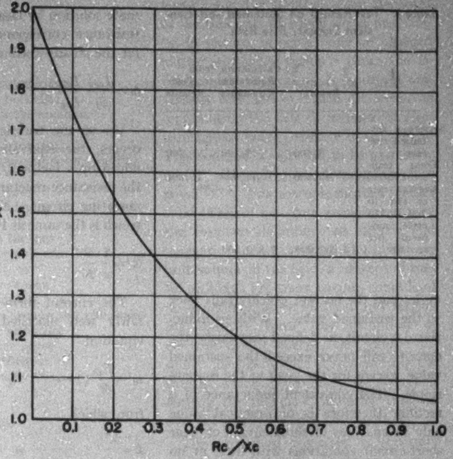


Fig. 2. Transient factor for determining crest value of short-circuit d-c transient current

Equivalent Machine Constants

Sustained Currents

For steady-state conditions, any part of the characteristic can be matched with an approximate equivalent resistance

$$\mathcal{R}=\frac{\Delta E_d}{\Delta I_d}=\frac{\Delta E_{dpu}}{\Delta S}\frac{E_{do}}{I_{dn}}X_{cpu} \quad (9)$$

The straight portions f and i of the net regulation curves for inductive and noninductive d-c circuits respectively can be represented in network equations by

$$\mathcal{R}_f=\frac{1}{2}\frac{E_{do}}{I_{dn}}X_{cpu}+mR_a+\frac{P_r}{I_{dn}^2} \quad (10)$$

and

$$\mathcal{R}_i=\frac{\pi}{4\sqrt{3}}\frac{E_{do}}{I_{dn}}X_{cpu}+mR_a+\frac{P_r}{I_{dn}^2} \quad (11)$$

The equivalent induced voltage in this region is

$$\mathcal{E}_0=E_{do}-E_{ao} \quad (12)$$

In many cases sufficient accuracy over the entire range is obtained by assuming a straight-line characteristic between limiting conditions. An average line, such as the one drawn to S_k for curve fgh in Fig. 1 is a good first approximation and will indicate absolute short-circuit current on the high side.

For greater accuracy, the regulation curve is represented by two straight lines, the first corresponding to curves f or i, and the second line drawn through S_k or S_j to match the lower section of the curve as closely as possible.

The slope of this second line represents the equivalent resistance for faults near short circuit. The intersection of the projection of this line with the zero ordinate determines the equivalent emf, $\mathcal{E}_s$, to be used in solving for near short-circuit currents.

If the solution leads to a current outside the range of the straight-line section chosen, the work must be repeated with another set of constants suitable to that region.

Transient Currents

If the d-c circuit is inductive, that is, $L_d>20X_c/2\pi f$, for the double-way circuit, or $L_d>5X_c/2\pi f$ for the double-wye, the equivalent internal inductance of the rectifier can be neglected for practical cases.

In any case if the external impedances involved in the fault are such that they will limit the short-circuit current during the first cycle to less than 0.55 I_{dn}/X_{cpu}, the equivalent internal inductance of the rectifier will be

$$\mathcal{L}=2X_c/2\pi f \quad (13)$$

for the double-way circuit, and

$$\mathcal{L}=1/2\ X_c/2\pi f \quad (14)$$

for the double-wye. For most purposes these inductances are negligible.

The report on protection[1] points out that a short-circuited rectifier reaches the transient crest current in less than 1

Table I. Test Results on Sustained and Transient Current, Rise Rate

	By Calculation, from		
	Report[1]	Equivalent Constants	Test Results
Average sustained current	8,500	8,100	7,700
Crest transient current	12,500	11,900	11,000
Average rate of rise in first 1/3 cycle, amperes per second	2,000,000	1,870,000	

cycle, and the current quickly falls back to the sustained value. With an inductive d-c circuit, as defined previously, the current will never exceed the sustained value, except for the peak of the 6-phase ripple. The equivalent inductance of a rectifier therefore is of practical value only in analyses of short-circuit or near short-circuit conditions with little or no inductance in the fault path.

The expedient of substituting two or more straight lines for actual characteristic of the rectifier in the solution of the transient condition has not been investigated. A step-by-step process is possible, but appears cumbersome.

For a first trial we can get an approximate solution by assuming an equivalent resistance corresponding to E_{do} and S_j for the steady-state conditions.

$$\mathcal{R} \approx \frac{X_{cpu}}{S_j} \frac{E_{do}}{I_{dn}} \tag{15}$$

Up to the transient peak, if a peak occurs, the equivalent resistance is reduced by a factor $3/(\pi K)$, dependent on the resistance-reactance ratio of the commutating circuit. K is given in Fig. 2, which is the same as Fig. 11 of reference 1.

$$\mathcal{R}' \approx \frac{3}{\pi} \frac{\mathcal{R}}{K} \tag{16}$$

The current form up to the peak is fairly well matched by the empirical equation

$$i_d' = \frac{E_{do}}{\mathcal{R}'}(1 - \epsilon^{-\omega t}) \tag{17}$$

from which

$$\mathcal{L} = \frac{\mathcal{R}'}{\omega} \tag{18}$$

Equations 17 and 18 tend to give currents on the high side in the first third of the cycle immediately after the fault is applied, and slightly low at the peak. Since $\mathcal{L}$ is generally very small, any appreciable external inductance will make the error negligible.

Conclusion

The equivalent constants presented for use in solving steady-state problems introduce errors which are understood and can be evaluated. The constants derived for transient analyses are less satisfactory but will provide usable answers if judiciously applied.

No check of the method by actual test has been made. However, curves *r*, *fgh*, and *ij* in Fig. 1 are based on the test rectifier described in Fig. 15 of the report on protection[1] and used for the short-circuit test of Appendix II of the report. Since only one point is available for comparison the construction of the net regulation curve is confirmed only at short circuit. See Table I.

References

1. PROTECTION OF ELECTRONIC POWER CONVERTERS, AIEE Committee Report. *AIEE Transactions*, vol. 69, pt. II, 1950, pp. 813–29.

2. POOL CATHODE MERCURY-ARC POWER CONVERTERS. *ASA C34.1—1949*, American Standards Association, New York, N. Y., 1949.

3. EXTENDED REGULATION CURVES FOR SIX-PHASE DOUBLE-WAY AND DOUBLE-WYE RECTIFIERS, I. K. Dortort. *AIEE Transactions*, vol. 72, pt. I, May 1953, pp. 192–202.

4. INFLUENCE OF A-C REACTANCE ON VOLTAGE REGULATION OF 6-PHASE RECTIFIERS, R. L. Witzke, J. V. Kresser, J. K. Dillard. *AIEE Transactions*, vol. 72, pt. I, July 1953, pp. 47–56.

No Discussion

Flow of Energy in D-C Machines

EDWARD I. HAWTHORNE
MEMBER AIEE

Synopsis: The flow of electromagnetic energy is traced in a typical, somewhat idealized, d-c generator, using an adaptation of Poynting's vector, after a discussion justifying this procedure. Quantitative analysis of the flow within the conductors and in the gap is carried out following the specification of the electromagnetic fields. Qualitative analysis of the flow in intermediate regions and to machine output leads is presented. Extension to other types of machines and to less idealized situations is discussed.

THE STUDY of energy flow in d-c machines, carried out on the basis of the electromagnetic field analysis, is presented. In the customary analysis of the machine the much simpler electric and magnetic circuit approach is employed as an excellent and useful approximation. Because of the complexity of the field approach to the analysis of electric devices at low frequencies in general, relatively few authors have considered problems in electric machinery on this basis.[1–4] The study of electromagnetic fields and energy flow in rotating machines affords a better understanding of their operation from an educational and analytical standpoint, and may be of value in improving their design. With this objective in mind this paper is presented.

Energy and Energy Flow Postulates

Before proceeding with the analysis of the flow of energy in a particular device such as the d-c machine, it is essential to ascertain clearly the basis for and the validity of the concepts of electromagnetic energy and energy flow.[5–9] Unnecessary confusion and concern over apparent paradoxes[10–11] frequently result from the failure to state clearly the fundamental aspects and limitations of these concepts.

Neglecting such nonlinear effects as saturation and hysteresis phenomena, as well as any problems requiring microscopic and relativistic considerations, all physical situations of electromechanical nature may be described completely by supplementing the relations of mechanics with the field equations of Maxwell and the force equation of Lorentz

$$\text{curl}\,\mathbf{E} = -\partial\mathbf{B}/\partial t;\ \text{div}\,\mathbf{D} = \rho \tag{1}$$

$$\text{curl}\,\mathbf{H} = \mathbf{J}_C + \rho v + \frac{\partial D}{\partial t};\ \text{div}\,\mathbf{B} = 0 \tag{2}$$

Paper 53-266, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 26, 1953; made available for printing May 8, 1953.

EDWARD I. HAWTHORNE is with the University of Pennsylvania, Philadelphia, Pa.

The author is deeply grateful to Dr. J. G. Brainerd of the University of Pennsylvania for his inspiring help and guidance in the study of the problem.

Nomenclature

Rationalized meter-kilogram-second units have been presumed.

a = number of paths in the armature
B = magnetic flux density
c = machine constant depending on load, equation 38
D = electric displacement density
E = electric field intensity
F = force
g = gap thickness
h = conductor height
H = magnetic field intensity
I = current
J (J_C) = (conduction) current density
k = machine parameter, equation 39
m = machine constant, equation 41
n = speed in revolutions per minute
p = number of poles
P = energy flow vector
q = electric charge
Q = number of slots, or commutator bars
r = radial co-ordinate
R = radius of the armature
s = thickness of conductors; distance
S = surface
t = time
T = normalized energy flow vector
u = pole arc
v = velocity
V = Maxwellian scalar potential
V_t = terminal voltage of the machine
w = pole pitch
x = circumferential co-ordinate
y = vertical co-ordinate in conductors (radial)
z = axial co-ordinate
Z = number of conductors
α = machine constant depending on load, equation 15
β = space angle of **P** in the gap
γ = number of slots per pole arc
δ = reciprocal of γ
ϵ_0 = permittivity of free space
ζ = normalized z co-ordinate
θ = angular co-ordinate around the armature
λ = axial length of the armature
μ_0 = permeability of free space
ξ = normalized x co-ordinate
ρ = electric charge density
σ = conductivity
τ = volume
ϕ = air gap flux per pole

$$d\mathbf{F} = dq(\mathbf{E}+\mathbf{v}\times\mathbf{B}) \qquad (3)$$

together with the empirical relations between the field vectors **D** and **E**, **B**, and **H**, and **J** and **E**, determined by the parameters of the medium. Maxwell's equations define these useful field vector concepts in consistence with experimental evidence, and the force equation provides for the essential tie with mechanics, in accordance with experimental fact. The only new postulated concept, which permits the description of electrical phenomena, not possible with the science of mechanics alone, is that of an electric charge, or current, or any other single quantity. The electromagnetic field formalism, defined by the field vectors of Maxwell, obviates the necessity of thinking in terms of action at a distance in favor of first determining the field of sources of electromagnetism, and then, employing this field concept with the aid of the Lorentzian force law, to study the effects as a local interaction between the field and charged matter. This procedure generally simplifies analysis both mathematically and conceptually.[12]

The role that the concept of electromagnetic energy plays in this formulation is not always made clear. The well-known fact[7,8,11] should be emphasized that there is nothing inherently present in Maxwell's equations which uniquely specifies such concepts as electromagnetic energy, electromagnetic energy storage density in the field, or electromagnetic energy flow vectors. These concepts are definitions based on the process of integration of the field equations. Carrying out the integration over a region of space, there results, after some manipulation, the energy balance equation

$$\underset{(a)}{\iiint \operatorname{div}(\mathbf{E}\times\mathbf{H})d\tau} +$$

$$\underset{(b)}{\iiint (\mathbf{E}\cdot\partial\mathbf{D}/\partial t+\mathbf{H}\cdot\partial\mathbf{B}/\partial t)d\tau} +$$

$$\underset{(c)}{\iiint \mathbf{J}_c\cdot\mathbf{E}d\tau}+\underset{(d)}{\iiint \rho\mathbf{v}\cdot\mathbf{E}d\tau}=0 \qquad (4)$$

By reference to the force equation 3 the terms c and d in equation 4 are identified as the time-rate at which work, mechanical in nature, is being done on charged matter within the region τ (since the **B**-field can do no work). It is almost universally accepted to regard the term b as representing the rate of increase of electromagnetic energy storage in the region. This concept has proved to be extremely useful, although other formulations are possible.[7,8] For isotropic media a further subdivision of b is made, which results in the common definitions of electric and magnetic energy storage densities per unit volume in the field, a concept of useful although arbitrary nature, capable of no experimental verification and requiring none.[7,8] With this definition of electromagnetic energy, the term a acquires the significance of representing the rate at which electromagnetic energy is leaving the region τ, in order that an over-all energy balance be preserved.

The procedure described is generally adapted to define the Poynting's energy flow vector at every point in space

$$\mathbf{P}=\mathbf{E}\times\mathbf{H} \qquad (5)$$

as representing the time-rate at which electromagnetic energy is transported through space per unit area. This definition, originally due to Poynting,[13] is again quite arbitrary, although extremely convenient, especially at higher frequencies. It must be realized that only a closed surface integral of **P**, to which the term a in equation 4 is reducible, has a true significance of over-all rate of flow of electromagnetic energy out of the enclosed region, with the conventional electromagnetic energy definition. Any other flow vector, differing from **P** by a curl of an arbitrary vector field, that is, having the same divergence as **P**, is equally satisfactory and significant.[9] Furthermore, if alternate postulates for electromagnetic energy are adopted, then other energy flow vectors, differing from **P** even in divergence, may be chosen for an equally valid formulation of the energy

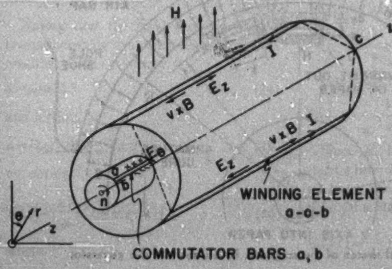
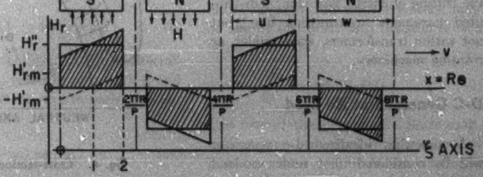


Fig. 1 (left). Typical winding element of a d-c generator

Fig. 2 (below). Radial magnetic field intensity versus distance around the gap



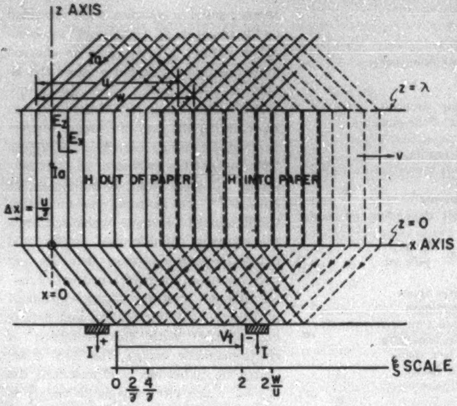


Fig. 3. Developed lap winding of the d-c generator

flow.[9] Thus it is not surprising that a literal application of the Poynting's vector leads to paradoxical occurrences of circulatory energy flow in cases where apparently no energy is being transmitted or transformed.[10–11] In all such cases, however, integration of the **P** vector over a closed surface invariably leads to an unambiguous resolution of the difficulty.

Space does not permit further discussion of the different aspects and various possibilities of flow vectors. The relative simplicity of Poynting's vector is ample justification for its selection in the analysis of the d-c machine to follow, with the following modifications. In the light of the foregoing, components of Poynting's vector having zero or negligibly small divergence may be omitted from it as representing an irrelevant circulatory flow of energy, in comparison with the components which represent a net transfer of power.

One such modification of the Poynting's vector leads, in its integrated form, to the commonly employed voltage-current postulate for power measurement. The connection between the voltage-current postulate and Poynting's vector has been investigated by Slepian[2,3] and others.[14] Thus the voltage-current approximation for the Poynting's vector is essentially valid in lumped parameter circuit studies, for the tracing of energy flow along a distributed parameter transmission line,[8] but not within transformers, wave guides, or rotating machinery.

D-C Generator at No Load

A somewhat idealized d-c generator will be considered, first under no-load conditions, in steady-state operation at constant speed. High-permeability iron with negligible saturation and hysteresis properties will be assumed, so shaped that the radial magnetic field in the gap is uniform under the pole shoes at no load and zero in the interpolar region of the gap. The effect of the presence of the large number of slots and of the ventilating ducts on the field in the gap will be neglected. Linear commutation will be assumed with the brushes located on the neutral axes. A 2-layer, simplex lap armature winding, with one turn per winding element, will be assumed, so that there are two conductors per slot, and the number of slots is equal to the number of commutator bars. To simplify the analysis, a permanent-magnet field excitation will be assumed. To study the flow of energy associated with the more common d-c field winding, an analysis of the flow of energy in a reactor is required. This has been worked out but will not be presented here, since the resulting modification of the analysis to follow would be of rather minor nature.

In order that the net force on the free charges within each winding element revolving in the steady magnetic field of the gap be zero in the steady state, an axial electric field must exist within and near the conductors on the surface of the armature, equal and opposite to the $\mathbf{v}\times\mathbf{B}$ term of equation 3. This field, an electrostatic one with respect to the fixed frame of reference, is associated with a charge distribution along the conductors and produces a potential difference between the commutator bars. Referring to Fig. 1, showing a typical winding element

$$E_z = -[\mathbf{v}\times\mu_0\mathbf{H}]_z \tag{6}$$

where

$$v = R\frac{d\theta}{dt} = \frac{\pi R n}{30} = \frac{wpn}{60} \tag{7}$$

Since the line integral of **E** around any closed path on the surface of the armature must be zero for the static **E**-field, there is also present an E_θ component everywhere, that is, between the charged armature conductors. A complete specification of this field is presented for the more general case of the loaded machine. One manifestation of this E_θ field is the voltage induced between commutator bars such as a and b in Fig. 1, given by

$$V_a - V_b = V_{ab} = \int_a^b E_\theta dx = \oint_{acb} E_z ds$$
$$= -\oint_{acb} \mathbf{v}\times\mathbf{B}\cdot d\mathbf{s} \tag{8}$$

The interaction between the **E** field and the **H** field in the gap results in a Poynting's vector whose divergence, however, is zero everywhere, as proved in the Appendix. The resulting apparent circulatory flow of energy may be

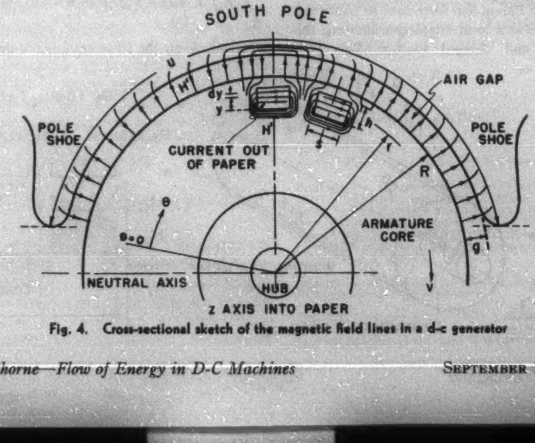


Fig. 4. Cross-sectional sketch of the magnetic field lines in a d-c generator

disregarded, and no net transfer of energy takes place.

Electromagnetic Field in Gap of a Loaded Generator

The simplest procedure to follow is to determine first the complete radial magnetic field in the gap, to which armature currents now contribute. Using equation 6, the electric field component E_z may be specified. Then the E_θ component may be determined by means of equation 8, or by the requirement that curl **E** be zero everywhere.

The radial magnetic field in the gap is shown developed in Fig. 2 for an idealized machine. Consider the developed plane of the gap, with the co-ordinates normalized to an axial co-ordinate ζ and circumferential co-ordinate ξ, given in the following:

$$\xi=2x/u;\ \zeta=(z/\lambda)(1/\gamma) \tag{9}$$

where

$$x=R\theta;\ \gamma=\left(\frac{u}{2\pi R}\right)Q \tag{10}$$

Let H_r'' be the constant radial field component caused by the field excitation and let H_r' be the component caused by armature currents. Its maximum value H_{rm}' can be calculated, considering the armature as a solenoid carrying a current I/a and having $(Z/p)(u/w)$ effective turns per pole, as follows

$$H_{rm}'=\frac{1}{2g}\,\frac{I}{a}\,\frac{Z}{p}\,\frac{u}{w}=\frac{I\gamma}{ga} \tag{11}$$

since $Z=2Q$ and $wp=2\pi R$. Taking as the origin for ξ the leading south pole tip, we have, for the region $0\leqslant\xi\leqslant 2$ under the south pole

$$H_r(\xi)=H_r''+H_{rm}'(\xi-1) \tag{12}$$

The electric field component $E_z(\zeta)$ in the gap can be computed directly from equations 6 and 12 if the conductors are assumed lossless. In the region $0\leqslant\xi\leqslant 2$, it is given by

$$E_z(\xi)=\mu_0 vH_r''+\mu_0 vH_{rm}'(\xi-1) \tag{13}$$

Noting that the total flux per pole ϕ is unaffected by armature reaction in this machine and is given by

$$\phi=\mu_0 u\lambda H_r'' \tag{14}$$

and letting

$$\alpha=H_{rm}'/H_r'' \tag{15}$$

we obtain

$$E_z(\xi)=\frac{aV_t}{2p\lambda\gamma}[1+\alpha(\xi-1)] \tag{16}$$

where

Table I. Axially Directed Power Flow Through the Boundary of the Air-Gap (Surface Integrals of the P_z Component)

	Left Half-Plane $0\leq\xi\leq 1$	Right Half-Plane $1\leq\xi\leq 2$	Total
Upper Boundary, $\zeta=1/\gamma$	$\frac{V_tI}{4p}(\alpha+\gamma-2\alpha\gamma/3)$	$\frac{V_tI}{4p}(-\alpha-\gamma-2\alpha\gamma/3)$	$\frac{V_tI}{4p}(-4\alpha\gamma/3)$
Lower Boundary, $\zeta=0$	$\frac{V_tI}{4p}(-\gamma+2\alpha\gamma/3)$	$\frac{V_tI}{4p}(\gamma+2\alpha\gamma/3)$	$\frac{V_tI}{4p}(4\alpha\gamma/3)$
Total	$\frac{V_tI\alpha}{4p}$	$-\frac{V_tI\alpha}{4p}$	0

$$V_t=\frac{np\phi}{60a} \tag{17}$$

For the simplex lap winding $p=a$, and hence

$$E_z(\xi)=\frac{V_t}{2\lambda\gamma}(1-\alpha+\alpha\xi) \tag{18}$$

The θ or x-directed component E_x can be determined with the aid of the developed lap winding shown in Fig. 3. The $E_z(\xi)$ component sets up a potential gradient in the (ξ,ζ) plane, which results in an E_x component that varies both with ξ and with ζ. Assuming the commutator bars to be closely spaced, E_x may first be determined at the $\zeta=0$ boundary directly from equation 8

$$\frac{2\pi R}{Q}E_x(\xi,0)=2\lambda E_z(\xi) \tag{19}$$

yielding

$$E_x(\xi,0)=\frac{V_t}{u}(1-\alpha+\alpha\xi) \tag{20}$$

Integrating equation 20 under a pole shoe over $0\leqslant\xi\leqslant 2$, that is, $0\leqslant x\leqslant u$, we obtain V_t, which can therefore be identified as the terminal voltage between brushes. Since equation 17 agrees with the usual expression for the terminal voltage of a lossless d-c machine, a check on the procedure has been obtained. At the boundary $\zeta=1/\gamma\ (z=\lambda)$ the following condition must be satisfied

$$E_x(\xi,1/\gamma)=E_x(\xi+\delta,0) \tag{21}$$

where δ is half the distance between successive commutator bars and is given in units of ξ as

$$\delta=\frac{1}{2}\,\frac{2\pi R}{Q}\,\frac{2}{u}=\frac{1}{\gamma} \tag{22}$$

Consequently

$$E_x(\xi,1/\gamma)=\frac{V_t}{u}(1-\alpha+\alpha/\gamma+\alpha\xi) \tag{23}$$

Since E_z is constant along each constant ξ line, E_x varies linearly with ζ, and we have, therefore, on the basis of equations 20 and 21

$$E_x(\xi,\zeta)=\frac{V_t}{u}(1-\alpha+\alpha\xi+\alpha\zeta) \tag{24}$$

Equation 24 could have been derived directly from equation 18 and from the condition curl $\mathbf{E}=0$, which condition equations 18 and 24 may now readily be seen to satisfy.

Had the losses within armature conductors not been neglected, equation 6 would have to be modified to read

$$E_z=-[\mathbf{v}\times\mu\mathbf{H}+\mathbf{J}/\sigma]_z \tag{25}$$

The expressions for E_z and E_x would, by a simple revision of proceeding steps, then become respectively

$$E_z=\frac{V_t}{2\lambda\gamma}(1-\alpha+\alpha\xi)-J_z/\sigma \tag{26}$$

and

$$E_x=\frac{V}{u}(1-\alpha+\alpha\xi+\alpha\zeta)-\frac{\lambda\gamma^2}{u}\,\frac{J_z}{\sigma}\zeta \tag{27}$$

Note that V_t, still given by equation 17, no longer represents the terminal voltage of the generator.

Energy Flow within Armature Conductors

The lines of the H' field are sketched in Fig. 4. Within the conductors, assumed for simplicity to be rectangular, the field H is x-directed and is given very nearly by

$$H_x(y)=-\frac{I}{a}\,\frac{y}{hs}=-Jy \tag{28}$$

The energy flow vector within the conductors is therefore directed radially everywhere and is given, using equations 5, 26, and 28, as

$$P_y(\xi,y)+\frac{V+J}{2\lambda\gamma}(1-\alpha+\alpha\xi)y-(J^2/\sigma)y \tag{29}$$

The divergence of this **P** vector, which represents the net generation of power per unit volume within the conductors, may be readily identified with the sum of two terms

$$\operatorname{div}\mathbf{P}=\frac{\partial Py}{\partial y}=\mathbf{v}\times\mathbf{B}\cdot\mathbf{J}-J^2/\sigma \tag{30}$$

The first of these represents electromagnetic power converted from mechanical power per unit volume of the conductor

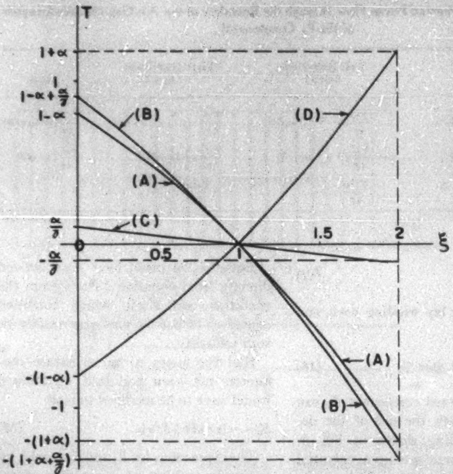


Fig. 5. Variation of the normalized energy flow vectors around the gap

Curve A: T_z(normalized P_z) at $z=0$
Curve B: T_z (normalized P_z) at $z=\lambda$
Curve C: Difference curve, (B)−(A)
Curve D: T_x(normalized P_x)

and the second constitutes the corresponding power loss. It may also be ascertained readily that the energy balance equation 4 is satisfied for every portion of the conductors. It is also satisfied for the armature as a whole, since the summation over all the conductors of the integral of P_y over the conductor surfaces at $y=h$ can be shown to yield

$$\sum_{\text{all conductors}}\left\{\int\int\int P_y(\xi,h)dS_y\right\}=\frac{Znp\phi I}{60a}-I^2\frac{1}{4}\frac{Z}{\sigma}\frac{\lambda}{sh}\quad(31)$$

which represents the generated power less the armature resistance losses.

Having indicated how internal losses may be included, we shall neglect them for the sake of simplicity in the analysis to follow.

The consideration of energy flow in the immediate vicinity of armature conductors, as it enters the gap, is postponed temporarily to study the conditions within the gap on a somewhat idealized basis.

Energy Flow in the Gap

As a result of the interaction of the H_r field with the E_x and E_z field, two components of **P** exist, namely

$$P_x=E_zH_r \text{ and } P_z=-E_xH_r \quad(32)$$

It will suffice to investigate the flow of energy under a south pole of Fig. 3, since both **E** and **H** reverse under the north pole, leaving the **P** vector unchanged. Furthermore, it will suffice to consider the armature reaction part of H_r. Previous reasoning based on the Appendix indicates that the H_r'' field produces components of **P** whose divergence is zero, and which, therefore, represent a circulatory flow of energy of no physical value. Using the $H_{rm}'\ (\xi-1)$ part of equation 12 for H_r together with equations 18 and 24, equation 32 yields

$$P_x(\xi)=\frac{IVt}{p}\frac{1}{2\lambda g}(\xi-1)(1-\alpha+\alpha\xi)\quad(33)$$

$$P_z(\xi,\zeta)=\frac{IV_t}{p}\frac{Q}{2\pi Rg}(1-\xi)(1-\alpha+\alpha\xi+\alpha\zeta)\quad(34)$$

These expressions are consistent with the energy balance equation 4. Integration of the P_x and P_z components over the side walls and over the front and rear walls respectively of a portion of the volume of the gap of height g, length λ, and width of one slot-tooth space, may be shown readily to yield, at any ξ, a power of $(2\lambda I/p)E_z(\xi)$ watts. This is exactly equal to the power emerging from the two conductors in the slot below, calculated from equation 29 for $\sigma=\infty$.

Investigation of the P_z and P_x components gives an insight on the flow of energy in the gap and reveals certain fundamental differences between them. The P_z component is $+z$-directed in the right half-plane ($\xi>1$) and $-z$-directed in the left half-plane ($\xi<1$). Although its divergence is not zero, the P_z component, when integrated over the front and rear walls of the gap at $\zeta=0$ and $\zeta=1/\gamma$ respectively produces zero net outflowing power from the region under a pole. (The same conclusion may be reached by noting that the volume integral of $\partial P_z/\partial z$ under a pole is zero.) The results of carrying out the surface integrals are summarized in Table I, which shows the outflowing, z-directed power at the boundaries of the gap under a pole.

The z-directed flow of energy may be best pictured by plotting the variable part of P_z, given by the normalized flow vector

$$T_z(\xi,\zeta)=(1-\xi)(1-\alpha+\alpha\xi+\alpha\zeta)\quad(35)$$

as a family of ζ curves versus ξ. Typical curves at the two extreme values of ζ are shown in Fig. 5, where $T_z\ (\xi,0)$ is drawn to scale for $\alpha=0.2$ as curve A, and where the displacement of the $T_z(\xi,1/\gamma)$ curve B from it for an assumed $\gamma=20$ is exaggerated ten times for the sake of clarity. It is evident that most of the z-directed energy which enters the gap at $z=0$, leaves it at $z=\lambda$. The net amount retained within the gap at any ξ is measured by the difference between the B and

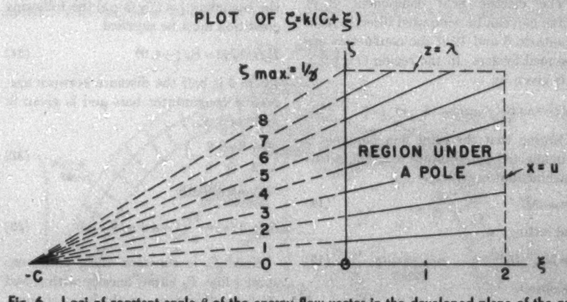


Fig. 6. Loci of constant angle β of the energy flow vector in the developed plane of the gap

0: k=0
1: $k=(1/8)k_{max}$
2: $k=(2/8)k_{max}$
.....
8: $k=k_{max}=1/\gamma C$

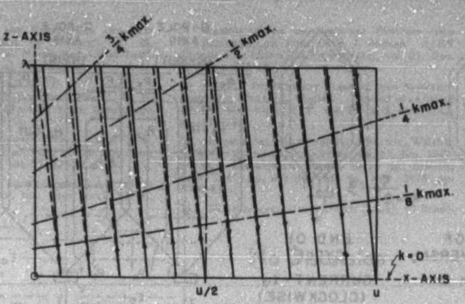


Fig. 7. Sketch of lines in energy flow in the gap. Dotted lines indicate conditions under light load

A curves and plotted as curve C in Fig. 5, also exaggerated ten times. This difference curve C is a linear function of ξ. The net power, measured by the area under this difference curve, is zero.

While the P_z component contributes markedly to the direction of the flow of energy at each point in the gap, it is the relatively smaller P_x component that accounts for the net transfer of power out of the gap. The normalized form of it

$$T_x(\xi)=(\xi-1)(1-\alpha+\alpha\xi) \tag{36}$$

which is independent of ζ, is also drawn in Fig. 5 for $\alpha=0.2$ as curve D. While the curves for T_x and T_z are of similar appearance, their meaning is quite different, since they are both plotted as functions of ξ. The plot of $T_x(\xi)$ indicates that energy flows in the $+x$ direction in the right half-plane and is $-x$ directed in the left half-plane. The amount of this outflowing energy is continuously increasing, as a result of the contributions from the upcoming P_y power from the armature conductors below, as the $\xi=0$ and $\xi=2$ boundaries are approached. The total outgoing power is given by the arithmetic sum of he lengths of the ordinates at $\xi=0$ and $\xi=2$ as

$$[(1-\alpha)+(1+\alpha)]\frac{V_tI}{2p\lambda g}(\lambda g)=\frac{V_tI}{p} \text{ watts per pole}$$

It is of interest to investigate the angle β which the lines of **P** make with the x-axis in the plane of the gap. It is given by

$$\tan\beta=P_z/P_x=-\frac{2\gamma\lambda}{u}\left(1+\frac{\zeta}{c+\xi}\right) \tag{37}$$

where

$$c=(1-\alpha)/\alpha \tag{38}$$

The angle β is constant along the family of straight lines given by

$$\zeta=k(c+\xi) \tag{39}$$

These constant $-\beta$ lines are sketched in Fig. 6 in the x-z plane for α-0.02, showing that the maximum value of $k=(\gamma c)^{-1} \ll 1$. In terms of k the angle β =

$$\beta=\pm\pi/2+\tan^{-1}\frac{m}{1+k} \tag{40}$$

where, for practical machines

$$m=\frac{u}{2\lambda\gamma}\ll 1 \tag{41}$$

Very nearly, then

$$\begin{aligned}\beta&\cong\pi/2+m-km,\ 0\leqslant\xi\leqslant 1\\ \beta&\cong-\pi/2+m-km,\ 1\leqslant\xi\leqslant 2\end{aligned} \tag{42}$$

The direction of the flow vector **P** in the gap is therefore very nearly constant in practical machines, except for a reversal of 180 degrees at each half-pole, and is almost parallel to the axis of the shaft. The largest difference in direction between the $0,\gamma$ and $u,0$ points in the x,z plane, is

$$|\Delta\beta|_{\max}=mk_{\max}=\frac{u}{2c\lambda\gamma^2} \tag{43}$$

which for $\alpha=0.2$ and $\gamma=20$ is only 0.0179 (u/λ) degrees. The largest deviation of β from the z-direction in this case is only 0.143 (u/λ) degrees. At light loads the lines of **P** make everywhere an angle of m radians with the z axis. As the load increases the **P** vector tends to curve towards the z-direction, while $|\Delta\beta|_{\max}$ increases.

The magnitude of **P** is given by

$$P=P_x\sqrt{1+\left(\frac{1+k}{m}\right)^2}\cong P_x\left(\frac{1+k}{m}\right) \tag{44}$$

It is very nearly independent of ζ and varies with ξ in the manner shown in Fig. 5(D) for T_x. The phase reversal occurs at $\xi=1$ where P diminishes to zero. Its variation with ζ along any constant ξ-line is caused by the variation of the k parameter, as illustrated in Fig. 6.

The lines P in the (x,z) plane are sketched in Fig. 7 for an exaggerated case of $k_{\max}=0.5$, $m=0.1$, corresponding to $\alpha=1/3$, $u/\lambda=1/5$, $\gamma=1$. Even for this much exaggerated case the angles are all very nearly equal to $\pi/2$. The lines of energy flow under light load are dotted for comparison, while the dashed lines represent the constant = β (or constant = k) lines of Fig. 6. Here the maximum angle with the vertical is 5.72 degrees and the maximum variation of this angle is 2.86 degrees.

Energy Flow from Armature Conductors to Gap

The transition, in which radially flowing energy, described by the P_y vector of equation 29, becomes the x and z

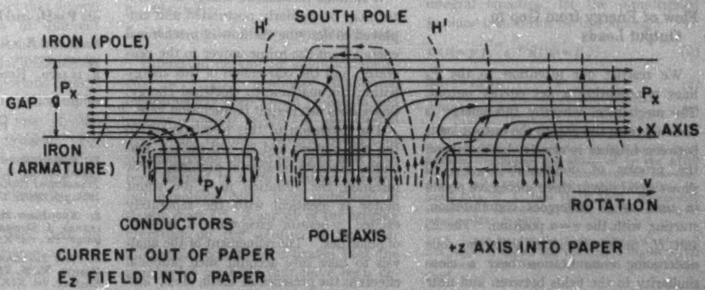
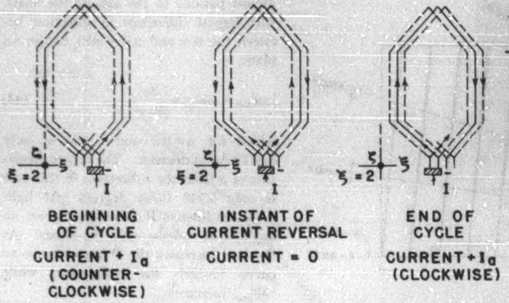


Fig. 8. Lines of H and P near armature conductors and in the gap, sectional view



BEGINNING OF CYCLE
CURRENT + I_a (COUNTER-CLOCKWISE)

INSTANT OF CURRENT REVERSAL
CURRENT = 0

END OF CYCLE
CURRENT + I_a (CLOCKWISE)

Fig. 9. Typical coil side undergoing commutation at successive instances

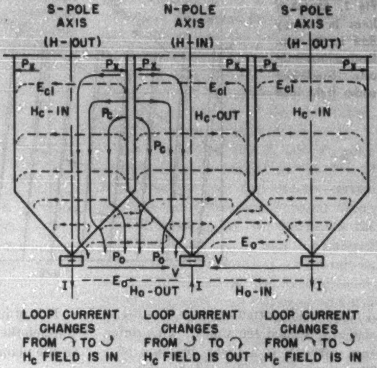

S-POLE AXIS (H-OUT)
N-POLE AXIS (H-IN)
S-POLE AXIS (H-OUT)
Px
Ecl
Hc-IN
Hc-OUT
Pc
Eo
Po
V
I
Ho-OUT
Ho-IN
LOOP CURRENT CHANGES FROM ↷ TO ↶ Hc FIELD IS IN
LOOP CURRENT CHANGES FROM ↶ TO ↷ Hc FIELD IS OUT
LOOP CURRENT CHANGES FROM ↷ TO ↶ Hc FIELD IS IN

Fig. 10. The flow of energy to the output terminals

directed flow of the P_x and P_z vectors of equations 33 and 34, can be studied if the magnetic field is specified, a very difficult task from the analytical standpoint. Qualitatively, the situation may be described with reference to Fig. 8, where lines of the H-field are sketched. It is seen how the lines of P_y gradually curve to become the lines of the P_x flow, as a result of the interaction with the E_z-field. Within the iron core, where H is small, there is very little flow of energy, except for the amount necessary to supply the stored energy and internal losses. The lines of E_z are not shown. The E_x field between the conductors in the iron teeth also contributes to the small energy flow there. In the gap the E_z field interacts with the H_r'-field to give the lines of energy flow their z-component, not shown. This interaction does not take place until the region well into the gap is reached where H_r' is appreciable. This curvature is out of the paper under the trailing half of each pole and into the paper under the leading half, as demonstrated in the preceding section. South pole is shown in Fig. 8 but, as discussed previously, the situation is unchanged under the north pole.

Flow of Energy from Gap to Output Leads

We restrict our attention to the P_x lines representing a net energy output. The mechanism whereby this energy is directed from the gap to the output leads between brushes is closely dependent on the process of commutation. Fig. 9 shows a coil side at three successive points in time as it undergoes commutation, starting with the $x=u$ position. The E_c and H_c fields associated with the coils undergoing commutation bear a close similarity to the fields between and near the wires of a transmission line, discussed by Slepian.[2] Space does not permit further elaboration here. Suffice it to say that the interaction of the E_{c1} field, produced by the charges along the conductors with the magnetic field H_c of the time-varying commutation currents, results in the flow vector P_c, whose lines are sketched on Fig. 10. The energy associated with the P_x vector enters the region of the neutral axis of the machine and is redirected by this P_c vector towards the output terminals, as shown on the sketch. Similarly the fields E_0 and H_0, associated with the output loop, result in the P_0 flow vector, which guides the energy out along the output transmission line of the generator. In an analogous manner the P_z-directed energy in the gap may be shown to be redirected to the adjacent regions of the gap by the leakage fields of the end turns, in accordance with the energy balance requirements of Table I.

Concluding Remarks

A suitable mechanical energy flow vector may be similarly postulated and employed to describe the flow of mechanical energy from the prime mover to the generator, and the conversion of this energy within the machine to electrical energy. Space does not permit its inclusion here.

The extension of this development to the d-c motors and to self-excited machines presents no special problem. For instance, the situation in a d-c shunt motor is identical with the separately excited generator except for a reversal of polarities. The refinement of the analysis to take into account such practical effects as the presence of teeth, slots, and ventilating ducts, nonuniform gap, or appreciably low permeability iron entails serious computational difficulties, and requires the exact specification of the fields, which must be obtained by a point-to-point field plot or by measurement. Eliminations of the assumption of linearity to include saturation and hysteresis effects presents a most serious computational as well as conceptual obstacle. This problem has not been attempted.

In spite of these limitations, it is hoped that the study of energy flow presented will serve the useful purpose of better basic understanding of the operation of d-c machines, and that it may be of value for designers of practical machines, which the somewhat idealized machine of this paper approaches.

Appendix

Let $\mathbf{E}_1$ be the electric field associated with a static charge distribution on the surface of a conductor 1 and let $\mathbf{H}_2$ be the magnetic field associated with a steady current distribution in a conductor 2. If $\mathbf{P}=\mathbf{E}_1\times\mathbf{H}_2$, then by equations 1 and 2

$$\operatorname{div}\mathbf{P}=\mathbf{H}_2\cdot\operatorname{curl}\mathbf{E}_1-\mathbf{E}_1\cdot\operatorname{curl}\mathbf{H}_2=-\mathbf{E}_1\cdot\mathbf{J}_2$$

since curl $\mathbf{E}_1=0$. However $\mathbf{J}_2=$ zero everywhere except within conductor 2, where $\mathbf{E}_1$ is zero. Hence div $\mathbf{P}=0$ everywhere.

References

1. Synchronous Machine with Solid Cylindrical Rotor, C. Concordia, H. Poritsky. *AIEE Transactions (Electrical Engineering)*, vol. 56, Jan. 1937, pp. 49–58, 179.

2. The Flow of Power in Electrical Machines, J. Slepian. *Electric Journal*, East Pittsburgh, Pa., vol. XVI, no. 5, 1919, pp. 303–11.

3. The Transformation of Energy in Power Plants, W. M. Thornton. *World Power*, London, England, vol. XIX, pt. 5, no. CIX, Jan. 1933, p. 13.

4. Foundations of the Electrical and Mechanical Transmission of Energy, W. M. Thornton. *Transactions*, North-East Coast Institution of Engineers and Shipbuilders, Newcastle-on-Tyne, England, vol. LVII, pt. 1; Nov. 1940.

5. The Theory of Electricity (book), G. H. Livens. Cambridge University Press, Cambridge, England, 1926, pp. 236–44.

6. The Electromagnetic Field (book), M. Mason, W. Weaver. University of Chicago Press, Chicago, Ill., 1929, pp. 264–72, 323–28, 1929.

7. Electromagnetic Theory (book), J. A. Stratton. McGraw-Hill Book Co., Inc., New York, N. Y., 1941.

8. Energy Flow in Electric Systems—the V-i Energy-Flow Postulate, J. Slepian. *AIEE Transactions (Electrical Engineering)*, vol. 61, Dec. 1942, pp. 835–41.

9. Energy and Energy Flow in the Electromagnetic Field, J. Slepian. *Journal of Applied Physics*, New York, N. Y., vol. 13, no. 8, Aug. 1942, pp. 512–18.

10. Electrical Essay for Recreation, Walther Richter. *Electrical Engineering*, vol. 66, Jan. 1947, p. 44. Discussion: March 1947, pp. 318–19; Sept. 1947, pp. 944–45; vol. 68, Nov. 1948, pp. 1130–31.

11. Electromagnetic Theory (book), S. Schelkunoff. D. Van Nostrand Company, Inc., New York, N. Y., 1943, pp. 77–81.

12. Fundamentals of Electrodynamics, W. F. G. Swann. *Bulletin*, National Research Council. Washington, D. C., vol. 4, pt. 6, no. 24, Dec. 1922, pp. 5–54.

13. On the Transfer of Energy in the Electromagnetic Field, J. H. Poynting. *Philosophical Transactions*, Royal Society of London, London, England, vol. 175, part II, 1884.

14. The Physics of a Transmission Line, W. M. Thornton. *Journal*, Institution of Electrical Engineers, London, England, vol. 88, pt. II, Dec. 1941, pp. 561–67.

No Discussion

Applications of Integral Equations to the Solution of Nonlinear Electric Circuit Problems

L. A. PIPES
MEMBER AIEE

Synopsis: This paper presents a mathematical technique useful in the solution of certain nonlinear electrical and mechanical problems of practical importance. The mathematical procedure involves the formulation of the problem in terms of nonlinear integral equations, and, by its use, practical results may be obtained in many cases where the approximate solution of the integral equation may be effected by the assistance of the Laplace transform theory. No attempt is made to present a general theory, but instead the general method is illustrated by applying it to the solution of typical problems of technical importance such as the analysis of electric circuits with nonlinear resistance, nonlinear inductance, nonlinear capacitance, and to the steady-state solution of van der Pol's equation.

THE MATHEMATICAL analysis of most of the phenomena that occur in nature usually leads to the formulation of the problems in terms of nonlinear differential equations. The fact that several of the classical theories of different branches of science have been formulated in terms of linear differential equations is evidence that in many important cases it is possible to linearize the equation under consideration without a loss of the essential features of the problem. For example, the theory of propagation of sound in air is based on the assumption that it involves adiabatic processes and that the pressure variation is infinitesimal, despite the fact that all audible sounds are associated with a finite pressure variation. The linear theory of sound explains the observed phenomena very well unless the sound intensity is great.

There are, however, certain phenomena that cannot be explained by a linearization of the equations that describe them, and it is then necessary to obtain a solution of the pertinent nonlinear differential equations in order to explain and predict the observed behaviour. At the present time, there is a growing importance in the study of nonlinear differential equations in connection with the following subjects:

1. Electrical circuits with saturable reactances and capacitances;
2. Large deflection elastic theory;
3. Mechanical vibrations;
4. Electrical machinery;
5. Aerodynamics;
6. Plasticity;
7. Wave motion of finite amplitude; etc.

Interest in developing analytical methods for solving nonlinear differential equations has been shown in Russia and is now growing greatly in the United States and Britain. Several books that summarize the various methods that have been devised to solve nonlinear differential equations have been published in recent years.[1–6] A survey of the existing literature on the solution of nonlinear differential equations indicates that the number of methods that have been devised for the solutions of nonlinear problems that arise in practice is small. It is the purpose of this discussion to present a method for solving nonlinear differential equations that is based on the theory of nonlinear integral equations,[7] and to apply it to some problems of importance in electrical engineering.

A Series Circuit with a Nonlinear Element

The general features of the method to be presented are clearly demonstrated by applying it to a series circuit containing a general nonlinear element. Such a circuit is represented by Fig. 1. This circuit consists of a linear component whose operational impedance, $Z(D)$ (where $D=d/dt$, the time derivative operator) is the argument of the impedance operator Z, in series with a nonlinear component. It will be assumed that the potential drop across the nonlinear component is a function $F(i)$ of the current that flows through it.

The differential equation that expresses the potential balance of the circuit may be written in operational form

$$Z(D)i(t)+F[i(t)]=e(t) \tag{1}$$

where $i(t)$ is the current that flows in the circuit and $e(t)$ is the applied potential. In order to formulate equation 1 as an integral equation, let the p-multiplied Laplace transforms be introduced.[8]

$$\mathfrak{L}i(t)=p\int_0^\infty e^{-pt}i(t)dt=I(p) \tag{2}$$

$$\mathfrak{L}e(t)=E(p) \tag{3}$$

$$\mathfrak{L}F[i(t)]=G(p) \tag{4}$$

Paper 53-265, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted November 10, 1952; made available for printing May 6, 1953.

L. A. Pipes is with the University of California, Los Angeles, Calif., and the United States Naval Ordnance Test Station, Inyokern, China Lake, Calif.

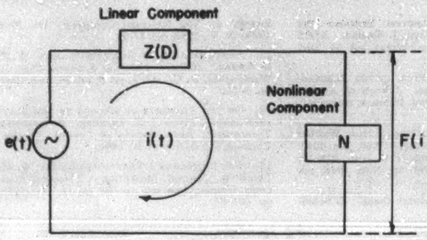


Fig. 1. General nonlinear circuit

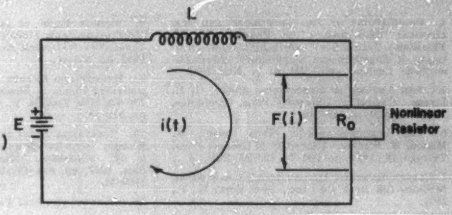


Fig. 2. Circuit with nonlinear resistor

Let it be assumed that initially the circuit is at rest so that at $t=0$, $i(0)=0$ and $q(0)=0$. That is, the initial current i and the charge separation, $q(t)$ of the circuit are both zero at $t=0$, therefore $i=\dot{q}$.

With these initial conditions, the Laplace transform of equation 1 is given by

$$Z(p)I(p)+G(p)=E(p) \tag{5}$$

or

$$I(p)=\frac{E(p)}{Z(p)}-\frac{G(p)}{Z(p)} \tag{6}$$

For simplicity, introduce the notation

$$H(p)=\frac{p}{Z(p)} \tag{7}$$

Hence equation 6 may be written in the form

$$I(p)=\frac{E(p)H(p)}{p}-\frac{G(p)H(p)}{p} \tag{8}$$

Let the inverse Laplace transform of $H(p)$ be $h(t)$, that is

$$\mathcal{L}^{-1}H(p)=\mathcal{L}^{-1}\frac{p}{Z(p)}=h(t) \tag{9}$$

The current of the circuit $i(t)$ may be obtained by taking the inverse Laplace transform of equation 8 in the form

$$i(t)=\mathcal{L}^{-1}\frac{E(p)H(p)}{p}-\mathcal{L}^{-1}\frac{G(p)H(p)}{p} \tag{10}$$

If the Faltung or superposition theorem of the theory of the Laplace transform[8] is applied to the right-hand member of equation 10, the result is obtained

$$i(t)=\int_0^t h(t-u)e(u)du-\int_0^t h(t-u)F[i(u)]du \tag{11}$$

If $F(i)=0$, so that the nonlinear component of the circuit is not present, then equation 11 reduces to

$$i_0(t)=\int_0^t h(t-u)e(u)du \tag{12}$$

This is the current that would flow in the circuit of Fig. 1 in the absence of the nonlinear component. Thus it is seen that equation 11 may be written

$$i(t)=i_0(t)-\int_0^t h(t-u)F[i(u)]du \tag{13}$$

This equation is known in the mathematical literature as Lalesco's nonlinear integral equation.[4] It can be shown[9] that the limit of the infinite sequence of functions

$$\begin{aligned}
i_0(t)&=\int_0^t h(t-u)e(u)du\\
i_1(t)&=i_0(t)-\int_0^t h(t-u)F[i_0(u)]du\\
i_2(t)&=i_0(t)-\int_0^t h(t-u)F[i_1(u)]du \qquad (14)\\
&\cdot\\
&\cdot\\
i_{n+1}(t)&=i_0(t)-\int_0^t h(t-u)F[i_n(u)]du
\end{aligned}$$

is the solution of integral equation 13. That is

$$i(t)=\operatorname*{limit}_{n\to\infty} i_n(t) \tag{15}$$

The investigation of the convergence of sequence 14 is given in reference 9. The results of this investigation indicate that the sequence converges rapidly in applications of this method of solution to special cases arising from physical problems.

Solution of the Integral Equation by an Operational Procedure

A convenient operational procedure which has been found useful in the calculation of the sequence of functions, $i_n(t)$, is based on the inverse Laplace transform of equation 6, which may be written in the form

$$i(t)=\mathcal{L}^{-1}I(p)=\mathcal{L}^{-1}\frac{E(p)}{Z(p)}-\mathcal{L}^{-1}\left\{\frac{\mathcal{L}F[i(t)]}{Z(p)}\right\} \tag{16}$$

The current $i_0(t)$ is given by

$$i_0(t)=\mathcal{L}^{-1}\frac{E(p)}{Z(p)} \tag{17}$$

Hence equation 16 may be written in the form

$$i(t)=i_0(t)-\mathcal{L}^{-1}\left\{\frac{\mathcal{L}F[i(t)]}{Z(p)}\right\} \tag{18}$$

The sequence of functions, $i_1(t)$, $i_2(t)$. . . $i_n(t)$ of equation 14 may be conveniently expressed by equations

$$\begin{aligned}
i_1(t)&=i_0(t)-\mathcal{L}^{-1}\left\{\frac{\mathcal{L}F}{Z(p)}[i_0(t)]\right\}\\
i_2(t)&=i_0(t)-\mathcal{L}^{-1}\left\{\frac{\mathcal{L}F}{Z(p)}[i_1(t)]\right\} \qquad (19)\\
&\cdot\\
&\cdot\\
i_{n+1}(t)&=i_0(t)-\mathcal{L}^{-1}\left\{\frac{\mathcal{L}F}{Z(p)}[i_n(t)]\right\}
\end{aligned}$$

In practice, it is usually found that the sequence, $i_1(t)$, $i_2(t)$, . . . may be more simply computed by equation 19 than by equation 14, because a table of suitable Laplace transforms assists in effecting the required integrations.

A Circuit with a Nonlinear Resistor

In order to illustrate the general method of the foregoing section, it will be applied to obtain the current in the circuit of Fig. 2. This circuit consists of a constant potential E in series with a linear inductor and a nonlinear resistor. It will be assumed that the potential drop across the nonlinear resistor is given by,

$$F(i)=R_0i^2 \tag{20}$$

where R_0 is a constant. If it is assumed that the potential E is suddenly impressed on the circuit at $t=0$, with zero initial current, then the current gradually increases until it reaches a final steady-state value. The differential equation of the circuit is

$$L\frac{di}{dt}+R_0i^2=E \quad \text{for } t>0 \tag{21}$$

In this simple case, the impedance $Z(p)=Lp$, where L is the inductance parameter of the linear inductor. The function $H(p)$ is now given by

$$H(p)=\frac{p}{Lp}=\frac{1}{L};\quad h(t)=\frac{1}{L} \tag{22}$$

In this case, integral equation 11 reduces to

$$t(i)=\int_0^t \frac{E}{L}dt-\frac{R_0}{L}\int_0^t i^3 dt \qquad (23)$$

The sequence of functions 14 may be obtained by direct integration, and are given by

$$i_0(t)=\frac{Et}{L}$$

$$i_1(t)=\frac{Et}{L}-R_0\frac{E^3t^3}{3L^3} \qquad (24)$$

$$i_2(t)=\frac{Et}{L}-R_0\frac{E^3t^3}{3L^3}+2R_0^2\frac{E^5t^5}{15L^5}-R_0\frac{E^4t^7}{63L^7}$$

$$\vdots$$

Differential equation 21 may be integrated by separating the variables, and writing it in the form

$$t=L\int_0^i \frac{di}{(E-R_0i^2)} \qquad (25)$$

If the indicated integration is performed, the current may be shown to be

$$i(t)=\sqrt{\frac{E}{R_0}}\tanh\left(\frac{t}{L}\sqrt{R_0E}\right) \qquad (26)$$

If equation 26 is expanded in powers of t, the correspondence between the sequence of functions, $i_1(t)$, $i_2(t)$,. . .and the power series expansion for $i(t)$ is clearly evident. The final steady-state current i_s is given by

$$i_s=\lim_{t\to\infty}\sqrt{\frac{E}{R_0}}\tanh\left(\frac{t}{L}\sqrt{R_0E}\right)=\sqrt{\frac{E}{R_0}} \qquad (27)$$

as is otherwise evident from physical considerations.

A Circuit with a Nonlinear Inductor

An important technical problem is the computation of the growth of current in a circuit containing a nonlinear inductor produced by the application of a constant potential source to a circuit such as the one depicted by Fig. 3. This circuit contains a constant potential source in series with a linear resistor and inductor. This combination is, in turn, in series with a nonlinear inductor. The nonlinear inductor consists of a coil of N turns, wound on a magnetic core having a cross-sectional area A and a mean length s. It will be assumed that at $t=0$, the circuit is inert and that the potential E_0 is impressed on the circuit at $t=0$, so that $i(0)=0$.

The differential equation of the circuit is

$$L\frac{di}{dt}+Ri+N\frac{d\phi}{dt}=E_0 \quad t>0 \qquad (28)$$

where ϕ is the magnetic flux in the core of the nonlinear inductor. In order to express the flux ϕ in terms of the current i, it is necessary to apply Ampère's law

$$Hs=Ni \qquad (29)$$

in suitable units, and to express the magnetization, or B—H curve of the magnetic material of the core in an analytical form. For many practical purposes, the magnetization curve may be represented by the third-order polynomial

$$B=\mu_0H-aH^3 \qquad (30)$$

where a is a constant determined empirically, and μ_0 is the initial permeability of the core material, defined by the equation

$$\mu_0=\left(\frac{dB}{dH}\right)_{H=0} \qquad (31)$$

If A is the mean cross-sectional area of the inductor core, the flux ϕ may be expressed in the form

$$\phi=BA=A(\mu_0H-aH^3)=\mu_0\frac{ANi}{s}-Aa\left(\frac{Ni}{s}\right)^3 \qquad (32)$$

It is useful to introduce the initial inductance of the nonlinear inductor, defined by the equation

$$L_0=\mu_0\frac{N^2}{s}A \qquad (33)$$

into the analysis. In terms of this parameter, equation 32 may be written in the form

$$\phi=L_0\frac{i}{N}-c_3i^3 \qquad (34)$$

where

$$c_3=Aa\frac{N^3}{s^3} \qquad (35)$$

If equation 34 is substituted into equation 28 the result is

$$(L+L_0)\frac{di}{dt}+Ri-Nc_3\frac{d\,i^3}{dt}=E_0 \qquad (36)$$

Let L_t be the total inductance of the circuit, given by

$$L_t=(L+L_0)=L+\mu_0\frac{N^2}{s}A \qquad (37)$$

and

$$k=Nc_3=Aa\frac{N^4}{s^3} \qquad (38)$$

In terms of these parameters, equation 36 may be written in the form

$$L_t\frac{di}{dt}+Ri-kDi^3=E_0 \qquad (39)$$

This equation is of the form of equation 1 with

$$Z(D)=L_tD+R \qquad (40)$$

and

$$F(i)=kDi^3 \qquad (41)$$

In this special case, we have

$$H(p)=\frac{p}{Z(p)}=\frac{p}{L_tp+R} \qquad (42)$$

$$h(t)=\mathfrak{L}^{-1}H(p)=\frac{\epsilon^{-bt}}{L_t}$$

where

$$b=\frac{R}{L_t} \qquad (43)$$

The integral equation satisfied by the current in the circuit, is in this case

$$i(t)=i_0(t)-\frac{k}{L_t}\int_0^t \exp[-b(t-u)]Di^3(u)du \qquad (44)$$

where

$$i_0(t)=\mathfrak{L}^{-1}\frac{E_0}{(L_tp+R)}=\frac{E_0}{R}(1-\epsilon^{-bt}) \qquad (45)$$

In order to obtain the approximate solution of nonlinear integral equation 44, it is convenient to write it in the operational form of equation 18; it then becomes

$$i(t)=\frac{E_0}{R}(1-\epsilon^{-bt})-\mathfrak{L}^{-1}\frac{[kLDi^3]}{(L_tp+R)} \qquad (46)$$

If equation 45 is cubed, the result is

$$i_0^3(t)=\frac{E_0^3}{R^3}(1-3\epsilon^{-bt}+3\epsilon^{-2bt}-\epsilon^{-3bt}) \qquad (47)$$

Hence by consulting a table of Laplace transforms,[8] we obtain

$$kLDi_0^3=\frac{kE_0^3}{R^3}p\left[1-\frac{3p}{(p+b)}+\frac{3p}{(p+2b)}-\frac{p}{(p+3b)}\right] \qquad (48)$$

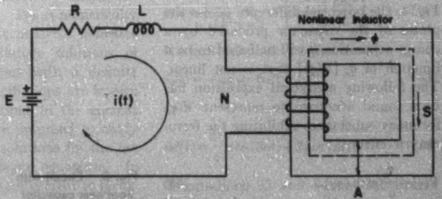


Fig. 3. Circuit with nonlinear inductor

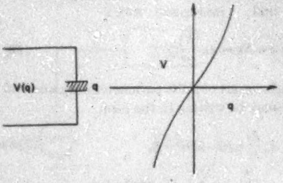


Fig. 4. Nonlinear capacitor and its saturation curve

The second approximation to the solution of equation 46 is given by equation 19; in this case it has the form

$$i_1(t)=\frac{E_0}{R}(1-\epsilon^{-bt})-\mathcal{L}^{-1}\frac{[kLDi_0{}^3]}{(L_tp+R)} \qquad (49)$$

By the use of equation 48 we obtain

$$\frac{kLDi_0{}^3}{(L_tp+R)}=\frac{kE_0{}^3}{L_tR^3}\left[\frac{p}{(p+b)}-\frac{3p^2}{(p+b)^2}+\frac{3p}{(p+2b)(p+b)}-\frac{p}{(p+3b)(p+b)}\right] \qquad (50)$$

The inverse transform of equation 50 is

$$\mathcal{L}^{-1}\frac{kLDi_0{}^3}{(L_tp+R)}=\frac{kE_0{}^3}{L_tR^3}\times\left[3bt\epsilon^{-bt}-\frac{9}{2}\epsilon^{-bt}+6\epsilon^{-2bt}-\frac{3}{2}\epsilon^{-3bt}\right] \qquad (51)$$

If equation 51 is substituted into equation 49 the second approximation for the current in the circuit is seen to be

$$i(t)=\frac{E_0}{R}(1-\epsilon^{-bt})-\frac{kE_0{}^3}{L_tR^3}\times\left[3bt\epsilon^{-bt}-\frac{9}{2}\epsilon^{-bt}+6\epsilon^{-2bt}-\frac{3}{2}\epsilon^{-3bt}\right] \qquad (52)$$

The higher approximations may be obtained by computing more functions of sequence 19.

A Circuit with a Nonlinear Capacitor

The dielectric properties of many substances have been discovered to have nonlinear characteristics as a consequence of the alignment of the electric dipoles of their atomic ions.[10] These substances are said to exhibit the ferroelectric effect. If one of these substances is used as the dielectrical of a capacitor, as shown in Fig. 4, the potential difference across the plates of the capacitor produced by a charge separation q will be found to be a function of q, $[V(q)]$ that is not linear. The following analytical expression has been found adequate to represent $V(q)$ for many substances exhibiting the ferroelectric effect.[11]

$$V(q)=\frac{S_0}{a}\sinh(aq) \qquad (53)$$

In this expression S_0 is the initial elastance of the capacitor defined by the equation

$$S_0=\left(\frac{dV}{dq}\right)_{q=0} \qquad (54)$$

This is the reciprocal of the initial capacitance defined by

$$C_0=\left(\frac{dq}{dV}\right)_{q=0} \qquad (55)$$

The constant a in equation 53 is chosen in such a manner so that this expression will give the best fit to the empirical saturation curve $V(q)$ of the capacitor under consideration. It may be noted that the quantity a is a measure of the degree of the nonlinearity of the nonlinear capacitor. That is, if $a=0$, the capacitor is linear and $V(q)=S_0q$.

To illustrate the general procedure in the analysis of a circuit containing a nonlinear capacitor,[11] consider the circuit of Fig. 5. This circuit consists of a linear resistor in series with a nonlinear capacitor and a periodic electromotive force. If the potential drop across the nonlinear capacitor is expressed by equation 53, Kirchhoff's electromotive-force law, applied to this circuit leads to the differential equation

$$R\frac{dq}{dt}+\frac{S_0}{a}\sinh(aq)=E_m\sin(\omega t) \qquad (56)$$

where q is the charge separation on the plates of the capacitor. The current of the circuit is given by

$$i=\frac{dq}{dt} \qquad (57)$$

In order to completely determine the problem, it will be assumed that at $t=0$, the harmonic potential is impressed on the circuit, and that $q(0)=0$. Let

$$b=\frac{S_0}{R} \qquad (58)$$

Equation 56 may be written in the form

$$\frac{dq}{dt}=\frac{E_m}{R}\sin(\omega t)-\frac{S_0}{aR}\sinh(aq) \qquad (59)$$

If the term $(S_0q)/R$ is added to both members of equation 59, the result may be written in the form

$$\frac{dq}{dt}+bq=\frac{E_m}{R}\sin(\omega t)+\frac{b}{a}[aq-\sinh(aq)] \qquad (60)$$

By the methods of the series-circuit section, it can be shown that the charge separation $q(t)$ satisfies the integral equation

$$q(t)=\frac{E_m}{R}\int_0^t \exp[b(u-t)]\sin(\omega u)du+\frac{b}{a}\int_0^t \exp[b(u-t)]\{aq(u)-\sinh[aq(u)]\}du \qquad (61)$$

In order to obtain the approximate solution of equation 61, it is expedient to proceed by an operational method similar to that of equation 19. To do this, introduce the Laplace transform of the charge separation $q(t)$ by the equation

$$\mathcal{L}_q(t)=Q(p) \qquad (62)$$

Since $q(0)=0$, the Laplace transform of equation 60 is

$$(p+b)Q=\frac{E_m}{R}\frac{\omega p}{(p^2+\omega^2)}+\frac{b}{a}\mathcal{L}[aq-\sinh(aq)] \qquad (63)$$

The inverse transform of equation 63 gives $q(t)$ in the form

$$q(t)=q_0(t)+\mathcal{L}^{-1}\left\{\frac{b}{a}\frac{\mathcal{L}[aq-\sinh(aq)]}{(p+b)}\right\} \qquad (64)$$

where

$$q_0(t)=\mathcal{L}^{-1}\frac{\frac{E_m}{R}\omega p}{(p+b)(p^2+\omega^2)}=\frac{E_m}{R}\omega\left\{\frac{\epsilon^{-bt}}{(b^2+\omega^2)}+\frac{\sin(\omega t-\theta)}{\omega\sqrt{(b^2+\omega^2)}}\right\} \qquad (65)$$

and

$$\tan\theta=\frac{\omega}{b}$$

Here, $q_0(t)$ is the linear approximation to the function $q(t)$. The second approximation $q_1(t)$ is given by substituting $q_0(t)$ into the right-hand member of equation 64 and is

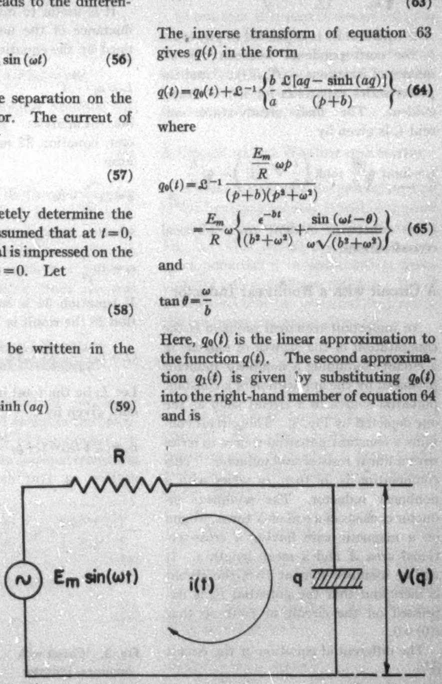


Fig. 5. Circuit with nonlinear capacitor

$$q_1(t)=q_0(t)+\mathcal{L}^{-1}\left\{\frac{b}{a}\mathcal{L}\left[\frac{aq_0-\sinh(aq_0)}{(p+b)}\right]\right\} \qquad (66)$$

The function $\sinh(aq_0)$ has the power series expansion

$$\sinh(aq_0)=(aq_0)+\frac{(aq_0)^3}{3!}+\frac{(aq_0)^5}{5!}+\frac{(aq_0)^7}{7!}+\ldots \qquad (67)$$

It has been found that for most nonlinear dielectric materials the parameter a is very small. If terms of the order of a^4 and higher are neglected

$$[aq_0-\sinh(aq_0)]\frac{1}{a}=-\frac{a^2}{6}q_0^3 \qquad (68)$$

To this order of approximation, equation 66 reduces to

$$q_1(t)=q_0(t)-\frac{a^2b}{6}\mathcal{L}^{-1}\left\{\frac{\mathcal{L}q_0^3}{(p+b)}\right\} \qquad (69)$$

THE STEADY-STATE SOLUTION

If the second order approximation to the steady-state response of the circuit is desired, the steady-state linear approximation $q_{0s}(t)$ after the transient term involving ϵ^{-bt} has disappeared may be used. The linear steady-state term is given by equation 65, in the form

$$q_{0s}(t)=A\sin(\phi) \qquad (70)$$

where

$$A=\frac{E_m}{R\sqrt{(b^2+\omega^2)}}$$

and

$$\phi=(\omega t-\theta) \qquad (71)$$

The second order steady-state approximation is given by substituting q_{0s} in the right-hand member of equation 69 in the form

$$q_{1s}(t)=q_{0s}(t)-\frac{a^2b}{6}\mathcal{L}^{-1}\frac{\mathcal{L}q_{0s}^3}{(p+b)} \qquad (72)$$

and retaining the periodic terms. Now

$$q_{0s}^3=A^3\sin^3\phi=\frac{A^3}{4}(3\sin\phi-\sin3\phi)=\frac{A^3}{4}[3\sin(\omega t-\theta)-\sin3(\omega t-\theta)] \qquad (73)$$

In order to compute the steady-state part of $q_{1s}(t)$ by the use of equation 72, let

$$f(t)=\mathcal{L}^{-1}\left[\frac{\mathcal{L}q_{0s}^3}{(p+b)}\right] \qquad (74)$$

Hence

$$\mathcal{L}f(t)=\frac{\mathcal{L}q_{0s}^3}{(p+b)} \qquad (75)$$

It can therefore be seen that $f(t)$ satisfies the differential equation

$$(D+b)f(t)=q_{0s}^3=\frac{A^3}{4}[3\sin(\omega t-\theta)-\sin(3\omega t-3\theta)] \qquad (76)$$

The periodic part of the solution of equation 76 is

$$f(t)=\frac{3A^3}{4}\frac{\sin(\omega t-2\theta)}{\sqrt{(b^2+\omega^2)}}-\frac{A^3}{4}\frac{\sin(3\omega t-3\theta-u)}{\sqrt{(b^2+9\omega^2)}} \qquad (77)$$

where

$$\tan(u)=\frac{3\omega}{b} \qquad (78)$$

Hence, the second order periodic or steady-state approximation as calculated from equation 72 is given by

$$q_{1s}(t)=A\sin(\omega t-\theta)-\frac{3a^2}{24}bA^3\frac{\sin(\omega t-2\theta)}{\sqrt{(b^2+\omega^2)}}+\frac{a^2}{24}bA^3\frac{\sin(3\omega t-3\theta-u)}{\sqrt{(b^2+9\omega^2)}} \qquad (79)$$

To this order of approximation, the steady-state charge oscillation is seen to contain a third-harmonic component not in phase with the fundamental. The steady-state current is given by

$$i_{1s}(t)=\frac{d}{dt}[q_{1s}(t)] \qquad (80)$$

Quasilinear Oscillations and the van der Pol Equation

A very important equation in the theory of vacuum tube oscillators is van der Pol's equation.[11] This equation is usually written in the form

$$\frac{d^2x}{dt^2}+\mu(x^2-1)\frac{dx}{dt}+\omega^2x=0 \qquad (81)$$

where μ and ω are constants.

The nature of the solutions of equation 81 for various values of the parameter μ has been exhaustively studied by van der Pol.[12,13] The case for μ a large number is the basis of van der Pol's theory of relaxation oscillations.[13] If $\mu=0$, the van der Pol equation reduces to that of a simple oscillator and the solutions of equation 81 are then of the form

$$x=C\sin(\omega t+\phi) \qquad (82)$$

where C and ϕ are arbitrary amplitude and phase constants. If μ is numerically small, van der Pol's analysis indicates that harmonic oscillatory solutions of equation 81 exist, but with a definite amplitude. Such oscillations are known as quasilinear vibrations in the current literature. It is here proposed to study the nature of these solutions by the use of the method of integral equations. In order to do this, it is convenient to recognize that the nonlinear term of equation 81 may be written in the form

$$x^2\dot{x}=\frac{1}{3}\frac{dx^3}{dt} \qquad (83)$$

In accordance with this, equation 81 may be written in the form

$$\frac{d^2x}{dt^2}+\omega^2x=\mu\left[\dot{x}-\frac{1}{3}\frac{dx^3}{dt}\right] \qquad (84)$$

Let it now be assumed that equation 84 satisfies the following conditions at $t=0$

$$x(0)=a;\quad \dot{x}(0)=0 \qquad (85)$$

Introduce the Laplace transform

$$\mathcal{L}x(t)=y(p) \qquad (86)$$

The transform of equation 84 with the initial conditions given by equations 85 is

$$(p^2+\omega^2)y=p^2a+\mu\mathcal{L}\left[\dot{x}-\frac{1}{3}\frac{dx^3}{dt}\right] \qquad (87)$$

Hence

$$y=\frac{p^2a}{(p^2+\omega^2)}+\frac{\mu}{(p^2+\omega^2)}\mathcal{L}\left[\dot{x}-\frac{1}{3}\frac{dx^3}{dt}\right] \qquad (88)$$

The inverse transform of equation 88 may be expressed in the form of the integral equation

$$x(t)=a\cos(\omega t)+\frac{\mu}{\omega}\int_0^t\sin[\omega(t-u)]\times\left\{\frac{dx}{du}(u)-\frac{1}{3}\frac{dx^3}{du}(u)\right\}du \qquad (89)$$

If $\mu=0$, the solution of the resulting linear equation is given by

$$x_0(t)=a\cos(\omega t) \qquad (90)$$

The second-order periodic solution of equation 81 may be obtained in the following manner, by a sequence similar to that of equation 19. Let the transform of the solution of linear equation 90 be given by

$$y_0=\mathcal{L}x_0(t)=\frac{p^2a}{(p^2+\omega^2)} \qquad (91)$$

Let $y_1(p)$ be the transform of the second approximation to the solution of equation 81. This is given by substituting $x_0(t)$ in the right-hand member of equation 88 in the form

$$y_1=y_0+\frac{\mu}{(p^2+\omega^2)}\mathcal{L}\left[\dot{x}_0-\frac{1}{3}\frac{dx_0^3}{dt}\right] \qquad (92)$$

We now calculate the functions

$$\dot{x}_0(t)=-\omega a\sin(\omega t) \qquad (93)$$

and

$$x_0^3(t)=\frac{a^3}{4}[\cos(3\omega t)+3\cos(\omega t)] \qquad (94)$$

If equations 93 and 94 are substituted into equation 92, the result is

$$y_1=y_0+\frac{\mu\omega}{(p^2+\omega^2)}+\frac{\mu\omega}{(p^2+\omega^2)}\mathfrak{L}\left[\left(\frac{a^3}{4}-a\right)\times \sin(\omega t)+\frac{a^3}{4}\sin(3\omega t)\right] \quad (95)$$

The indicated Laplace transforms may be obtained from a table of transforms and equation 95, written as

$$y_1=y_0+\mu\left(\frac{a^3}{4}-a\right)\frac{p\omega^2}{(p^2+\omega^2)^2}+\frac{3}{4}\mu a^3\omega^2\frac{p}{[p^2+(9\omega)^2](p^2+\omega^2)} \quad (96)$$

A table of Laplace transforms yields the result

$$\mathfrak{L}^{-1}\frac{p\omega^2}{(p^2+\omega^2)^2}=\frac{1}{2\omega}[\sin(\omega t)-(\omega t)\cos(\omega t)] \quad (97)$$

Thus, it is seen that unless

$$\left(\frac{a^3}{4}-a\right)=0 \quad (98)$$

the inverse transform of equation 96 will give rise to a secular term of the form of equation 97, and the solution will not be periodic. Hence the possible amplitudes a of the periodic steady-state solution to this order of approximation are the solutions of equation 98. These are given by

$$a=0 \quad \text{or} \quad a=\pm 2 \quad (99)$$

This is the well-known result for the possible amplitudes of the periodic solutions of equation 81, for small values of μ obtained by van der Pol[12] using another method of analysis. If $a=2$, the second-order periodic solution is obtained by computing the inverse transform of equation 96. This solution is

$$x_1(t)=2\cos(\omega t)+\frac{\mu}{4\omega}[3\sin(\omega t)-\sin(3\omega t)] \quad (100)$$

For small values of μ/ω, the first term of the right member of equation 100 is the dominant one. This is the quasilinear solution of equation 81.

Conclusion

The mathematical technique for the solution of certain nonlinear differential equations of importance in electric circuit theory has as its theoretical basis the theory of Lalesco's nonlinear differential equation.[7] It has been found, however, that in order to obtain practical results in an efficient manner, an operational procedure such as that given in the series-circuit section is preferable to the classical iteration procedure given by equation 14.

Application of the methods given in this paper to the solution of several practical nonlinear problems has indicated that useful results can be obtained with an economy of effort. The method is particularly efficient in the solution of differential equations in which the nonlinear terms are multiplied by parameters whose numerical values are small.

References

1. Introduction to Nonlinear Mechanics (book), **N. Minorsky.** Edwards Brothers Publishers, Ann Arbor, Michigan, 1947.

2. Theory of Oscillations (book), **A. A. Andronov, C. E. Chaikin.** Princeton University Press, Princeton, N. J., 1949.

3. Introduction to Nonlinear Mechanics (book), **N. Kryloff, N. Bogoliuboff.** Princeton University Press, Princeton, N. J., 1943.

4. Nonlinear Vibrations (book), **J. J. Stoker** Interscience Publishers, New York, N. Y., 1950.

5. Nonlinear Differential Equations (book), **N. W. McLachlan.** Clarendon Press, Oxford, 1950.

6. The Engineer Grapples with Nonlinear Problems, **T. von Kármán.** *Bulletin*, American Mathematical Society, New York, N. Y., vol. 46, 1940, p. 615.

7. Leçons sur les Équationes Intégrales (book), **Vito Volterra.** Gauthier-Villars, Paris, France, 1913.

8. Applied Mathematics for Engineers and Physicists (book), **L. A. Pipes.** McGraw-Hill Book Company, Inc., New York, N. Y., 1946.

9. Introduction à la Théorie des Équationes Intégrales (book), **T. Lalesco.** Gauthier-Villars, Paris, France, 1912.

10. Polar Molecules, **P. Debye.** Dover Publications, New York, N. Y., 1950.

11. A Mathematical Analysis of a Dielectric Amplifier, **L. A. Pipes.** *Journal of Applied Physics*, New York, N. Y., vol. 23, no. 8, Aug. 1952, pp. 818–24.

12. Nonlinear Theory of Electric Oscillations, **B. van der Pol.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 22, 1934, p. 1051.

13. Relaxation Oscillations, **B. van der Pol.** *Philosophical Magazine*, London, England, vol. 2, 1926, p. 928.

No Discussion

The First-Order Behavior of Separable Oscillators

DAVID C. dePACKH
NONMEMBER AIEE

Synopsis: The first-order performance of oscillators satisfying Groszkowski's criterion of separability of linear and nonlinear elements can be predicted in the first order with a knowledge only of the Nyquist diagram of the linear portion of the circuit, the amplitude characteristic of the nonlinear portion, and the loop gain. The best form of nonlinear characteristic from the standpoint of frequency stability is one in which the odd-power terms are dominant.

IT IS THE purpose of this paper to present the analysis of electric oscillators whose linear and nonlinear elements are separable in the first order, that is, under the assumption that the deviation of the oscillation from the simple harmonic form is small compared with the total amplitude. Although this criterion of separability is a stringent one in practice because of Miller effect, grid loading, plate resistance variations, and similar phenomena which essentially make the Nyquist[1] diagram a function of the amplitude, it is nevertheless clear that the principles of separable oscillators admit of a simple extension to the nonseparable case in the first order; the only additional requirement is a knowledge of the behavior of the resonant point of the Nyquist diagram with amplitude. This behavior can evidently be obtained with the feedback loop open. The frequency variation imposed by the separable oscillator theory is then added to this variation to give the total frequency behavior. In many practical cases the separability condition is nearly enough satisfied so that, except for the most precise results, the separable oscillator theory can be applied without correction.

This work will be termed a first-order analysis although in actuality second-order effects such as the frequency shift require the retention of terms of the

---

Paper 53-267, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted January 6, 1953; made available for printing May 4, 1953.

David C. dePackh is with the Naval Research Laboratory, Washington, D. C.

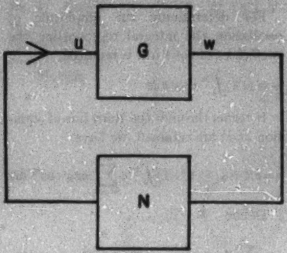


Fig. 1. Feedback loop of electric oscillator. Arrow indicates direction of energy flow

second order of small quantities. The analysis is evidently applicable also to oscillators of any nonelectrical, for example, mechanical, type which can be separated into linear and nonlinear elements whose functions are respectively expressed by a Nyquist diagram and a power series amplitude characteristic.

The problem of electric oscillators has been treated by many authors, and van der Pol[2] has shown how the amplitude may be derived in the first order with a knowledge of the characteristic of the nonlinear element. There is also considerable literature to indicate the effect of the harmonic content of the oscillation on the frequency of the system, at least for ordinary resonant inductance-capacitance circuits or their equivalent as the frequency-determining element. The literature, however, is lacking in a general treatment of the behavior of the system in terms of the nonlinear characteristic and a general form of frequency-determining element, and in general criteria for frequency stability with variation of the feedback. It is true that in the latter connection the knowledge of the frequency behavior in terms of the harmonic content has led certain workers, for example, Groszkowski,[3] to insert filters to remove the harmonics and in that way to arrive at a constant frequency. There are other means to stabilize the frequency by operating on the linear, frequency-determining portion of the circuit, either by making the harmonic responses of the network such that the total frequency shift contributed by the harmonic content vanishes, or by the use of a circuit whose effective Q is high, so that $\partial\psi/\partial\omega$ is large at resonance, where ω is the angular velocity and ψ is the phase on the Nyquist diagram. There is, however, an important criterion relative to the form of the nonlinear characteristic which apparently has not been explicitly given.

Reference 2 indicates by implication what this criterion must be. Van der Pol has indicated that the even-power nonlinear terms in the nonlinear characteristic are almost without effect in determining the amplitude since the power integral for such terms vanishes identically. One may conclude immediately from this observation that in any oscillating system in which even-power nonlinear terms are dominant the amplitude is likely to grow very large, to grow, in fact, until the odd-power content can limit the amplitude. But by this time the even-harmonic content is likely to be so large that the frequency will have undergone a considerable change. It would therefore appear that the best way of maintaining stable frequency behavior is to make the odd-power terms of the nonlinear characteristic dominant. An experimental verification of this conjecture as well as a theoretical one are given herein.

To treat this problem generally it seems appropriate to use the Nyquist diagram rather than to derive results for any specific system. It is plain that since this diagram can be experimentally obtained, and since the oscillating system is assumed to be in the steady state, there should be no need to have recourse to a differential equation.

Theoretical Data

The oscillating system is best thought of as comprising a nonlinear gain G and a frequency-determining linear element N in a closed loop, see Fig. 1. The functions of these two elements are assumed to be separate, which means that the gain G is single-valued in its output w with respect to its input u. Therefore the integral of $w\,du$ over 1 cycle of the oscillation must vanish; thus

$$\oint w\,du=0 \tag{1}$$

This integral was introduced in a very similar form by Groszkowski.[3] Now suppose

$$u\equiv A\cos\theta+\sum_{n=2}^{\infty}\epsilon_n\cos(n\theta+\phi_n)$$

and

$$w\equiv C[v_0+A'\cos(\theta+\phi)+\sum_{n=2}^{\infty}\epsilon_n'\cos(n\theta+\phi_n')]$$

where C is a constant. The element N will be assumed to have zero response at $\omega=0$, this being the case in nearly all practical circumstances. Hence u is to contain no constant term.

Then carrying out the integration, Equation 1 gives

$$AA'\sin\phi=-\sum_{n=2}^{\infty}n\epsilon_n\epsilon_n'\sin(\phi_n'-\phi_n)$$

θ is of course $\omega_0 t$, where ω_0 is the steady-state angular velocity. The amplitudes and phases of the various harmonics are related through the passive element N in the following way

$$\epsilon_n=C\epsilon_n'|N_n|;\ \phi_n-\phi_n'=\psi_n$$

where $N(\omega)$, the gain of the element N, is given by

$$N(\omega)=|N(\omega)|\epsilon^{j\psi(\omega)}$$

Accordingly

$$\sin\psi_1=-\sum_{n=2}^{\infty}n(\epsilon_n/A)^2(1/\nu_n)\sin\psi_n;$$

$$\nu_n\equiv|N_n|/|N_1| \tag{2}$$

Since ψ_1 corresponds to the altered fundamental frequency on the Nyquist diagram, equation 2 is analogous to the well-known expressions giving the frequency shift in terms of the harmonic content.

Let us now alter slightly the arrangement of Fig. 1 by multiplying the gain G and dividing the factor N by a constant equal to the scalar gain of N at resonance. This changes the passive element into N', which has unity gain at its resonant frequency. The element N' will be referred to hereafter as the Nyquist. The resulting nonlinear element G' may be divided into a linear factor k and a resid-

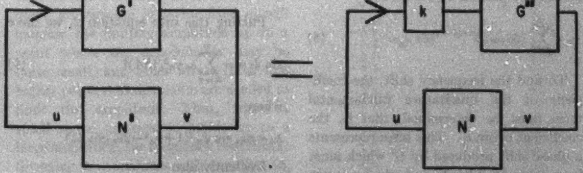


Fig. 2. Equivalent circuits to Fig. 1

ual distorting element G''; see Fig. 2. The characteristic of G' is then expressible as

$$v=ku+a_2(ku)^2+a_3(ku)^3+\dots \tag{3}$$

and we take $v=w/C$.

If $u=A\cos\theta+\epsilon$, where ϵ represents the harmonic content, we have

$$\begin{aligned}v=&kA\cos\theta+\\ &k\epsilon+\beta_2\cos^2\theta+\beta_3\cos^3\theta+\dots+\\ &(\epsilon/A)(2\beta_2\cos\theta+3\beta_3\cos^2\theta+\dots)+\\ &(\epsilon^2/A^2)(\beta_2+3\beta_3\cos\theta+6\beta_4\cos^2\theta+\dots)+\\ &\dots\end{aligned} \tag{3A}$$

where $\beta_m\equiv a_m\ (kA)^m$. The expression thus written is arranged in ascending order, since β_m is assumed to be a small quantity of the order of ϵ. Only the first three lines of the expression will be retained.

Before proceeding further, it is worth while to derive a relation of the form

$$\sum_{m=2}^{\infty}\beta_m\cos^m\theta=\sum_{n=0}^{\infty}\gamma_n\cos n\theta \tag{4}$$

One has

$$\begin{aligned}\cos^m\theta&=[(\epsilon^{j\theta}+\epsilon^{-j\theta})/2]^m\\ &=(1/2^m)\{(\epsilon^{jm\theta}+\epsilon^{-jm\theta})+\\ &\quad m[\epsilon^{j(m-2)\theta}+\epsilon^{-j(m-2)\theta}]+\\ &\quad[m(m-2)/2!][\epsilon^{j(m-4)\theta}+\\ &\qquad\epsilon^{-j(m-4)\theta}]+\dots\}\\ &=(1/2^{m-1})[\cos m\theta+\\ &\qquad m\cos(m-2)\theta+\dots]\\ &=(1/2^{m-1})\sum_{k=0}^{[m/2]}B_{km}\cos(m-2k)\theta\end{aligned}$$

where $B_{km}\equiv m!/[k!(m-k)!]$ for $m\neq 2k$, $B_{(m/2),m}\equiv m!/[2(m/2)!^2]$, this case occurring only for m even.

Then

$$\sum_{m=2}^{\infty}\beta_m\cos^m\theta=\sum_{m=2}^{\infty}(\beta_m/2^{m-1})\sum_{k=0}^{[m/2]}B_{km}\times\cos(m-2k)\theta$$

On writing $n\equiv m-2k$ and rearranging the order of summation, one has

$$\sum_{m=2}^{\infty}\beta_m\cos^m\theta=\sum_{m=k}^{\infty}\sum_{k=0}^{\infty}(\beta_{n+2k}/2^{n+2k-1})\times B_{k,n+2k}\cos n\theta$$

$(\beta_0=\beta_1\equiv 0)$

Hence in equation 4 γ_n is given by

$$\gamma_n=\sum_{k=0}^{\infty}(\beta_{n+2k}/2^{n+2k-1})B_{k,n+2k} \tag{5}$$

To find the frequency shift, the coefficient of the quadrature fundamental term must be determined, that is, the coefficient of $\sin\theta$. This term represents a phase shift produced by G' which must be compensated by N' and which accordingly produces a change in frequency.

If σ is the coefficient of $\sin\theta$, then

$$\sigma=(1/\pi)\int_0^{2\pi}v\sin\theta\,d\theta$$

Evidently only the third line of equation 3(A), neglecting all lines beyond the third, can contain $\sin\theta$. The value of the third line is

$$\begin{aligned}-(\epsilon/A\sin\theta)(d/d\theta)&\sum_{m=2}^{\infty}\beta_m\cos^m\theta\\ &=(\epsilon/A\sin\theta)\sum_{n=0}^{\infty}n\gamma_n\sin n\theta\end{aligned}$$

from equation 4. Therefore

$$\begin{aligned}\sigma&=(1/\pi A)\int_0^{2\pi}\left(\sum_{n=2}^{\infty}-\epsilon_n\sin n\theta\sin\phi_n\right)\times\left(\sum_{n=0}^{\infty}n\gamma_n\sin n\theta\right)d\theta\\ &=-(1/A)\sum_{n=2}^{\infty}n\epsilon_n\gamma_n\sin\phi_n\end{aligned} \tag{6}$$

Now $\sigma/kA=\tan\psi_1\cong\sin\psi_1$, so that equation 6 divided by kA must be equal (in the first order) term by term with equation 2. Therefore

$$\epsilon_n=(\nu_n\gamma_n/k)(\sin\phi_n/\sin\psi_n) \tag{7}$$

If only the first two lines of equation 3(A) are retained, it is found that

$$\begin{aligned}v&=ku+\sum_{n=0}^{\infty}\gamma_n\cos n\theta\\ &=\gamma_0+(kA+\gamma_1)\cos\theta+\\ &\qquad\sum_{n=2}^{\infty}[k\epsilon_n\cos(n\theta+\phi_n)+\gamma_n\cos n\theta]\\ &=v_0+A'\cos\theta+\sum_{n=2}^{\infty}\epsilon'_n\cos(n\theta+\phi'_n)\end{aligned}$$

neglecting, in the latter relation, the second-order fundamental phase shift. Hence we derive, on comparing the phase angles of the harmonics

$$\begin{aligned}\cot\phi'_n&=\cot(\phi_n-\psi_n)\\ &=\cot\phi_n+\gamma_n/(k\epsilon_n\sin\phi_n)\end{aligned}$$

Combining this with equation 7, we can solve it, finding

$$\cot\phi_n=\cot\psi_n-\nu_n\csc\psi_n \tag{8}$$

so that

$$\epsilon_n=(\gamma_n\nu_n/k)(1+\nu_n^2-2\nu_n\cos\psi_n)^{-1/2} \tag{7A}$$

Putting this into equation 2, we have

$$\sin\psi_1=-\sum_{n=2}^{\infty}\lambda_n\gamma_n^2/(kA)^2 \tag{9}$$

where

$$\lambda_n=n\nu_n\sin\psi_n/(1+\nu_n^2-2\nu_n\cos\psi_n)$$

Evidently also

$$\Delta\omega\cong-\sin\psi_1/(\partial\psi/\partial\omega)_{\psi=0} \tag{10}$$

For determining the amplitude of oscillation the integral representing the coefficient of $\cos\theta$ in v is required

$$\rho\equiv(1/\pi)\int_0^{2\pi}v\cos\theta\,d\theta$$

If terms through the third line of equation 3(A) are retained, we have

$$\rho=kA+\gamma_1+(1/\pi A)\int_0^{2\pi}\epsilon\sum_{m=2}^{\infty}m\beta_m\cos^m\theta d\theta$$

Define

$$\delta_n\equiv\sum_{k=0}^{\infty}[(n+2k)\beta_{n+2k}/2^{n+2k-1}]B_{k,n+2k} \tag{11}$$

then

$$\begin{aligned}\rho&=kA+\gamma_1+(1/\pi A)\int_0^{2\pi}\sum_{n=2}^{\infty}\epsilon_n\cos n\theta\times\cos\phi_n\sum_{n=0}^{\infty}\delta_n\cos n\theta\,d\theta\\ &=kA+\gamma_1+(1/A)\sum_{n=2}^{\infty}(\nu_n\gamma_n\delta_n/k)\times(\sin\phi_n\cos\phi_n/\sin\psi_n)\\ &=kA+\gamma_1+(1/A)\sum_{n=2}^{\infty}(\nu_n\gamma_n\delta_n/k)\times(\cos\psi_n-\nu_n)/(1+\nu_n^2-2\nu_n\cos\psi_n)\end{aligned}$$

using equations 7 and 8.

But since $\sin\psi_1$ is of the second order, it follows that $\cos\psi_1$ differs from unity by a quantity of the fourth order. Hence we may put $\rho=A$, and the equation determining the amplitude is thus

$$(k-1)A+\gamma_1+(1/kA)\sum_{n=2}^{\infty}\gamma_n\delta_n\mu_n=0 \tag{12}$$

to the second order, where

$$\mu_n=\nu_n(\cos\psi_n-\nu_n)/(1+\nu_n^2-2\nu_n\cos\psi_n)$$

It will be seen from equation 5 that γ_1 depends only on odd-order values of β_m and hence of a_m in equation 3. Therefore the first-order limiting is given by

$$\begin{aligned}(k-1)A+\gamma_1=(k-1)A+(3/4)a_3(kA)^3+\\ (5/8)a_5(kA)^5+\dots=0\end{aligned} \tag{13}$$

this being the relation derived by van der Pol[2] and others.

Evidently if only even-order distortion exists, $\gamma_1=0$, and there is no first-order limiting. Suppose

$$v=ku+a_2(ku)^2$$

Then $\gamma_2=(a_2/2)(kA)^2$, $\delta=a_2(kA)^2$, the other ordered values of γ_n and δ_n being zero.[3] Then the equation determining the limiting is

$$k(k-1)A^2+(1/2)a_2^2(kA)^4\mu_2=0$$

giving as the nontrivial solution

$$A=(1/k)[-2(k-1)/(k\mu_2a_2^2)]^{1/2} \tag{14}$$

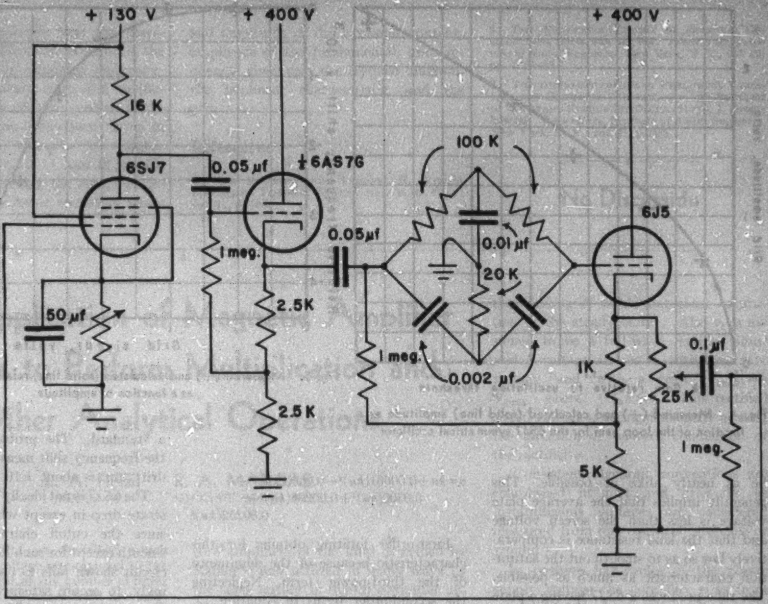


Fig. 3. Oscillator circuit showing nominal component values of Nyquist

Thus there can be second-order limiting only for $\mu_2<0$, that is, for $\cos\psi_2<\nu_2$. If this condition is not satisfied, there is no limiting in this case. Next suppose

$$v=ku+a_2(ku)^2+a_3(ku)^3$$

This yields the equation

$$(3/16)a_3^2\mu_3(kA)^4+[(3/4)a_3+(1/2)\mu_2a_2^2]\times (kA)^2+(1/k)(k-1)=0$$

for which the solution is

$$(kA)^2=-2[(3a_3+2\mu_2a_2^2)/(3\mu_3a_3^2)]\times \{1-[1-3\mu_3a_3^2(k-1)/k(3a_3+2\mu_2a_2^2)^2]^{1/2}\}$$

If the second term in the radical is small compared with unity, which is true if $k\cong 1$, this solution reduces to

$$A=(2/k)[(k-1)/(-3a_3-2\mu_2a_2^2)]^{1/2} \quad (15)$$

It may happen in cases of strong second-power distortion that the term $-3a_3$ is not sufficiently positive to make the radicand of equation 15 positive. In this case the first-order theory is no longer adequate to predict the behavior.

The frequency behavior can be determined by equations 9 and 10 once the amplitude is known. For the case of combined second- and third-power distortion.

$$\sin\psi_1=-(1/4)\lambda_2a_2^2(kA)^2-(1/16)\lambda_3a_3^2(kA)^4 \quad (16)$$

The second harmonic term is seen to be the worst in disturbing the frequency since A is proportional to the square root of the excess gain $(k-1)$, so that the frequency shift varies linearly with the excess gain for second-power distortion but only with its square for third-power distortion. When any even- or odd-power distortion is present by itself, it will contribute a frequency shift proportional to the excess gain, if even, or to its square, if odd, as can be inferred from equations 5, 9, and 12.

The reason for the familiar phenomenon of "dropping in" of oscillators can be readily perceived from equations 15 and 16. In this behavior the transition of the system to the oscillatory state is abrupt and the frequency undergoes a marked decrease. If in equation 15 $\mu_2>0$ and is not small, corresponding to fairly strong response and small phase shift of the second harmonic in the Nyquist, an increase in a_2 for fixed a_3 will increase the limiting amplitude up to a point where the denominator may be quite small, and other forms (fifth and higher powers) of distortion are needed to limit the amplitude. Thus, when a small signal appears, it builds up to a large amplitude, the frequency continually dropping in accordance with equation 16, until equilibrium is ultimately reached because of these higher power terms. This condition is probable in any oscillator whose characteristic has strong second-power distortion such as a triode, or a pentode in class AB, B, or C.

Experimental Data

An attempt was made to verify these predictions with a pentode oscillator operating with plate, screen, and bias voltages such that the cutoff and saturation portions of the characteristic would

Table I. Measured Static Characteristic of 6SJ7 as Connected in Figure 3

Grid Voltage (V_G)	Load Voltage (V_L)
−9.0	0.5
−8.5	1
−8.0	2
−7.5	4
−7.0	7
−6.5	11
−6.0	16.5
−5.5	23
−5.0	32
−4.5	42
−4.0	54
−3.5	65
−3.0	77
−2.5	90
−2.0	99
−1.5	105
−1.0	108
−0.5	110
0.0	112

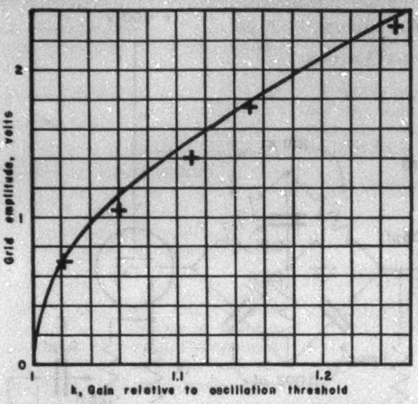


Fig. 4. Measured (+) and calculated (solid line) amplitude as a function of the loop gain for the 6SJ7 symmetrical oscillator

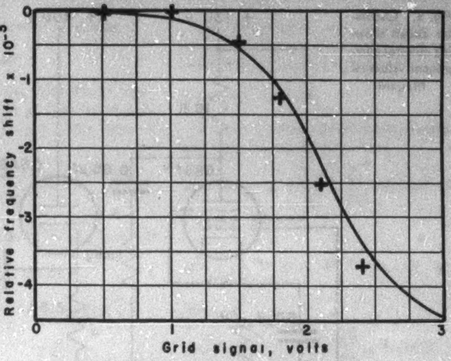


Fig. 5. Measured (+) and calculated (solid line) relative frequency shift as a function of amplitude

be as nearly alike as possible. This generally implies that the average plate voltage is less than the screen voltage and that the load resistance is comparatively low so as to smooth out the saturation characteristic as much as possible. The tube used was a *6SJ7* having a plate load of 16,000 ohms and 130 volts on the screen and as the plate supply voltage. A series cathode resistor was used to adjust the grid bias to the middle of the characteristic at -3.0 volts; see Fig. 3 and Table I.

Using a 7-point Lagrange formula based on $V_G=-3.0$ volts as the center with three 1.0-volt increments in each direction, the total range thus running from -6.0 volts to 0 volts for V_G, one finds

$$-V_L=77+23.95V_G'+0.1583V_G'^2-1.510V_G'^3-0.719V_G'^4+0.0688V_G'^5+0.0604V_G'^6$$

where $V_G'=V_G+3.0$. We take the normalized unit of amplitude as 1.0 volt, that is, $kA=1$ implies 1.0-volt signal at the grid. On dividing by the slope of the characteristic at the operating point, we put this relation into the terminology of this paper, thus

Table II. Measured Amplitude and Phase Response of Nyquist of Figure 3

Harmonic (n)	ν_n	$\sin\psi_n$	
2	2.1	−1.00	
3	3.3	−0.92	$-\psi_n>\pi/2$
4	4.2	−0.85	
5	4.7	−0.80	
6	5.0	−0.72	

$$v=ku+0.00661(ku)^2-0.0630(ku)^3-0.0300(ku)^4+0.00287(ku)^5+0.00252(ku)^6$$

First-order limiting obtains for this characteristic because of the dominance of the third-power term. Neglecting the second-order terms in equation 12, we have

$$kA=\{13.12-[172-557(k-1)/k]^{1/2}\}^{1/2}$$

A graph of this quantity, together with points obtained by measurement, is given in Fig. 4.

In order to obtain freedom from effects on the Nyquist produced by plate resistance changes with amplitude as well as by grid circuit loading which would result from initial velocity effects, etc., as the grid voltage approaches cathode potential, the two cathode followers (*6J5*, *1/2 6AS7G*) were used to obtain complete isolation of the frequency-determining element. The operating conditions of these are such as to introduce negligible amplitude distortion.

The resonant frequency for the nominal circuit elements both by computation and measurement is 790 cycles per second to sufficient accuracy; the relative responses and sines of the phase shifts for the first six harmonics are given in Table II.

The experimental and calculated values of $[\partial\psi/\partial(\omega/\omega_r)]_{\psi=0}$, ω_r being the resonant angular velocity, agree at approximately -3.7. These figures can be used to calculate the frequency shift by equations 5, 9, and 10. Fig. 5 shows the measured and calculated frequency shifts. Frequency departures were measured by timing beats between the oscillator and a standard. The probable accuracy of the frequency shift measurements due to drift, etc., is about $\pm 10^{-4}$.

The *6SJ7* is not ideally suited to demonstrate drop-in except when underbiased, since the cutoff characteristic is too smooth except for such low gains that the circuit shown fails to oscillate. Accordingly, to secure satisfactory even-power distortion a type *6AB7* was used in the same circuit. The inflexion point of the characteristic was found to be at -1.9 volts bias on the grid, and Fig. 6 shows the relative frequency shift for various values of bias with the circuit generating 1.0 volt of signal measured at the grid.

Conclusion

It is felt that there is adequate experimental and theoretical verification to

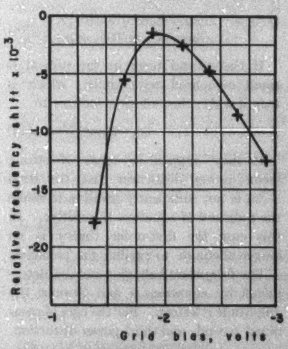


Fig. 6. Measured frequency behavior of 6AB7 oscillator showing effect of even-power distortion away from inflexion point

support the hypothesis that the dominant factor, apart from adjustments of the linear elements, in obtaining frequency-stable and smoothly controlled oscillations in separable electric systems is the suppression of the even-power terms in the amplitude characteristic. Equations 5, 7(A), 8, 9, 10, 11, and 12 provide means for determining the complete behavior of the oscillator, including phase and amplitude of the harmonic content, amplitude of the fundamental, and frequency, given only the Nyquist diagram, the nonlinear characteristic, and the gain.

References

1. THE REGENERATION THEORY, H. Nyquist. *Bell System Technical Journal*, New York, N. Y. vol. 11, p. 126.

2. THE NONLINEAR THEORY OF ELECTRIC OSCILLATIONS, Balth. van der Pol. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 22, pp. 1051–86.

3. THE INTERDEPENDENCE OF FREQUENCY VARIATION AND HARMONIC CONTENT, AND THE PROBLEM OF CONSTANT-FREQUENCY OSCILLATIONS, J. Groszkowski. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 21, p. 958.

No Discussion

An Application of Magnetic Amplifier Circuits to Perform Multiplication and Other Analytical Operations

L. A. FINZI
MEMBER AIEE

R. A. MATHIAS
ASSOCIATE MEMBER AIEE

Synopsis: Circuits are described which yield an output voltage proportional to the product or quotient of any number of input voltages. The components used are magnetic cores of grain-oriented nickel alloys, half-wave rectifiers, and linear resistors. Experimental results and some design considerations are given. Circuits of the same kind are also suggested for performing indefinite integration and other analytical operations upon signal voltages which are functions of time.

CONVENTIONAL magnetic amplifiers are well suited to perform mixing of signals as the output depends on the total magnetomotive force of control provided jointly by any number of input windings. Over the linear range of the ampere-turns transfer characteristic, the output is simply proportional to the algebraic summation of signals; other operations can also be performed, in a less straightforward manner, by proper use of nonlinear regions of the characteristic.

The circuits described in this paper, however, operate on somewhat different principles. Their components are single-core amplifiers of the kind described by Ramey[1,2] and utilized in many ways by the same author in digital computer applications.[3]

Although this kind of amplifier is known, a brief discussion of its essential features is given; a proper appreciation of these features is essential for the understanding of the specific applications described further in this paper.

Elementary Single-Core Amplifier

An elementary single-core amplifier is shown in its simplest circuitry in Fig. 1(A). The magnetic core is characterized by a sharply defined saturation at flux levels $\pm\Phi_{sat}$ reached with comparatively small magnetomotive forces, as shown in Fig. 1(B), in an idealization which is adequate for grain-oriented nickel alloys commonly used in high-grade amplifiers. A single winding links this core with N turns; r is the resistance of this winding and R_L is a load resistance. γ and σ are two half-wave rectifiers; it is assumed at first that forward rectifier conduction takes place with negligibly small voltage drops, while infinite resistance is offered to the flow of current in the reverse direction.

In Fig. 1(C), v_g is the voltage of the power supply, or gate voltage; v_s is the signal voltage. For the sake of simplicity in Fig. 1(C) it is assumed that $v_g = V_{g\max}\sin\omega t$, with $\omega = 2\pi f$ = angular velocity of the sinusoidal power source and $V_{g\max} < \omega N\Phi_{sat}$ (that is, application of the gate voltage directly at the terminals of the winding N would not cause saturation in the steady state). Also v_s is assumed to be a full-wave rectified sinusoidal voltage in phase with the gate voltage and with a crest $|V_{g\max}| = k$ $V_{g\max}$ where $0 \leq k \leq 1$. This rectified signal voltage is acting constantly against the direction of forward conduction of the rectifier σ.

(Consistent algebraic conventions are established in text and drawings as instantaneous voltages are considered measured from one common reference or ground.)

In positive half-cycles of the power supply voltage the core flux rises from some initial level towards saturation; in this process of varying fluxes the voltage v_g is balanced almost completely by the rate of change of flux linkages $Nd\Phi/dt$ while a very small magnetizing current (dictated by the minor hysteresis loop described) flows with negligibly small voltage drops in R_L and r. At some time during this half-cycle, saturation is reached; the process of flux changes terminates, and in the remaining part of this half-cycle $i = v_g/(R_L + r)$. As v_g becomes zero, i becomes zero, too, and at the end of the positive half-cycle the core is left at the flux level $+\Phi_{sat}$. No current flows in the signal circuit throughout this half-cycle because of the blocking action of the rectifier σ.

In the next (negative) half-cycle v_g reverses its polarity; the rectifier γ blocks any flow of current through the load. But, as v_g now overrules v_s, the rectifier σ becomes conducting and a process of flux changes initiates during which a reversed small magnetizing current flows in the winding. That is, the core flux is being brought down from the level $+\Phi_{sat}$, as the rate of change of flux linkages balances almost entirely the voltage $(v_g - v_s)$. In view of the limitation stipulated for the amplitude of v_g, in no case negative saturation is reached in this half-cycle, that is, the process of flux decay from $+\Phi_{sat}$ extends throughout this presetting half-cycle, at the end of

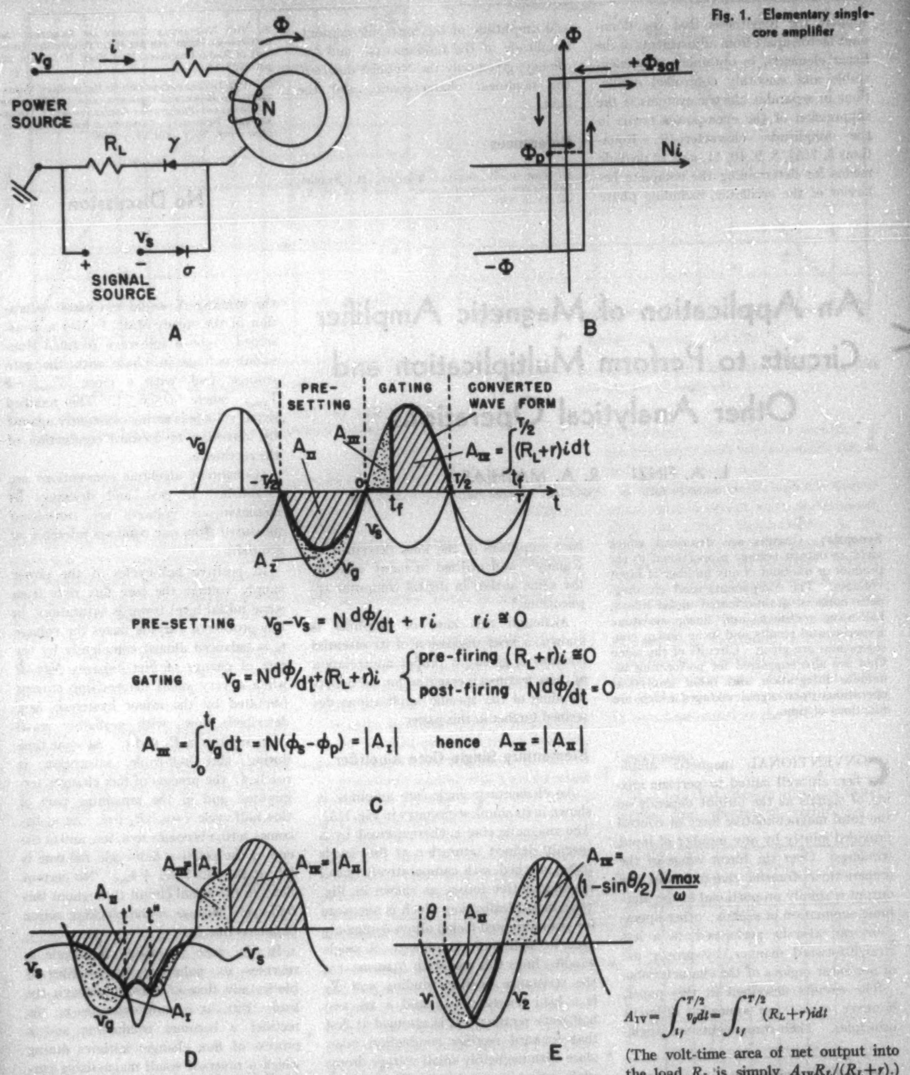
---

Paper 53-285, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted February 27, 1953; made available for printing May 6, 1953.

L. A. FINZI and R. A. MATHIAS are with the Carnegie Institute of Technology, Schenley Park, Pittsburgh, Pa.

This work was performed in part under sponsorship of the United States Office of Naval Research. The assistance of R. R. Jackson in the final experimentation is acknowledged.

Fig. 1. Elementary single-core amplifier


Φ
vg
i
r
N
POWER SOURCE
RL
γ
vs
+
−
σ
SIGNAL SOURCE
A
Φ
+Φsat
Φp
Ni
−Φ
B
PRE-SETTING
GATING
CONVERTED WAVE FORM
vg
−T/2
AII
AIII
0
AIV = ∫ T/2 tf (RL+r)i dt
tf
T/2
T
t
vs
AI
vg
PRE-SETTING vg − vs = N dϕ/dt + ri   ri ≅ 0
GATING vg = N dϕ/dt + (RL+r)i { pre-firing (RL+r)i ≅ 0; post-firing N dϕ/dt = 0
AIII = ∫ tf 0 vg dt = N(ϕs − ϕp) = |AI| hence AIV = |AII|
C
AIII=|AI|
AIV=|AII|
AII
t'
t"
vs
vs
vg
AI
D
AIV = (1 − sin θ/2) Vmax/ω
θ
AII
v1
v2
E

which a flux level Φ_p is reached, which is

$$-N(\Phi_{sat}-\Phi_p)=\int_{-T/2}^{0}(v_g-v_s)dt=\text{volt-time area } A_{\text{I}}$$

with the notation used in Fig. 1(C).

In the next gating half-cycle the flux starts rising again from this preset level Φ_p under the action of the positive v_g as described initially. This process terminates as a firing time t_f at which saturation is reached, that is

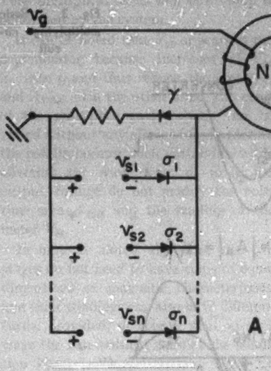
$$\int_0^{t_f} v_g dt = N(\Phi_{sat}-\Phi_p)=\text{volt-time area } A_{\text{III}}$$

From then on and for the rest of the gating half-cycle, the voltage v_g appears across the circuit resistances with a volt-time area of gross output

$$A_{\text{IV}}=\int_{t_f}^{T/2} v_g dt=\int_{t_f}^{T/2}(R_L+r)i dt$$

(The volt-time area of net output into the load R_L is simply $A_{\text{IV}}R_L/(R_L+r)$.) Since $A_{\text{III}}=-A_{\text{I}}$ it follows that $A_{\text{IV}}=-A_{\text{II}}$. Thus the controlling action of the signal voltage acting in one presetting half-cycle upon the output of the immediately following gating half-cycle is fully evidenced.

Although Fig. 1(C) refers to the simple case of a sinusoidal gate voltage v_g and a full-wave rectified sinusoidal signal volt-



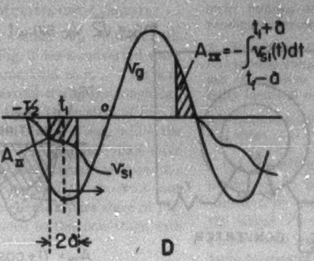
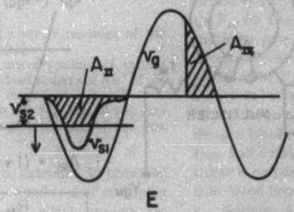
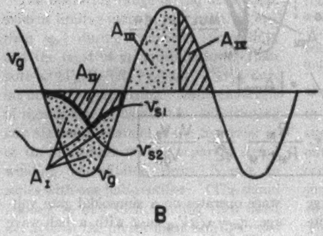
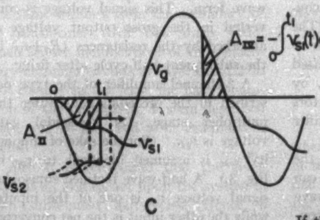


Fig. 2. Multichannel amplifier







age v_s, no use has been made of such particular restrictions throughout the reasoning. In fact, for the type of operation described, the gate voltage can have different wave forms and the durations of positive and negative half-cycles do not even need to be equal as long as the volt-time areas of v_g in one presetting half-cycle and in the subsequent gating half-cycle are equal in absolute value (and less than $2N\Phi_{sat}$).

It is seen also that wave form and magnitudes of signal voltage v_s during a gating half-cycle are of no relevance on the operation as long, of course, as σ is kept blocked. (That is, v_s should not be allowed to become positive but it can well be zero.)

During the presetting half-cycle, the wave form of signal may also be irrelevant. If the gate voltage is able to overrule v_s throughout this half-cycle, the fact remains that the spotted volt-time area A_{III} is equal and opposite to the time integral of $(v_g - v_s)$ over the presetting half-cycle. On the other hand, unduly large instantaneous values of negative v_s may cause σ to block, thus arresting the presetting process of flux changes temporarily. For instance, no flux changes take place over the interval $t't''$ in the situation of Fig. 1(D). In this case A_{II} no longer is the total volt-time area of signal, but rather is the shaded area of the contour of least ordinates (least in absolute value) of both v_s and v_g as shown.

This consideration allows predicting the performance of an amplifier with sinusoidal gate voltage and d-c signals, and explains its lack of linearity. (Linearity, of course, can be obtained in the case of d-c signals if the gate voltage is square-waved.) Also an elementary use of this consideration can be made for the measurement of the phase-angle θ of two sinusoidal voltages of equal amplitude and frequency $v_1 = V_{max} \sin \omega t$ and $v_2 = V_{max} \sin (\omega t - \theta)$. For this, v_g is made equal to v_1 while v_s is obtained from half-wave rectification of v_2. As seen in Fig. 1(E) the volt-time area A_{II} of the contour of leastordinates is then $(\sin \theta/2 - 1)\, V_{max}/\pi f$. Therefore, θ can be obtained from the readings of a D'Arsonval meter in the output circuit on a scale suitably marked.

These considerations relate volt-time areas of presetting to volt-time areas of gating. Thus the action of presetting signal substantially is related to cyclic averages of output currents or voltages. The wave form of the output voltage is also of interest in many instances; in the operation the wave form of the curve delimiting A_{II} is converted in the output into a wave form with ordinates equal to zero before firing and coinciding with the ordinates of the gate voltage after firing. This conversion of wave forms is utilized in one stage of the multiplying circuits described later in this paper.

Multichannel Amplifier

A modification of the previous device is shown in the amplifier of Fig. 2(A), in which many input channels are provided for a corresponding number of signal voltages $v_{s1}, v_{s2} \ldots v_{sn}$. In gating half-cycles the operation is the same as

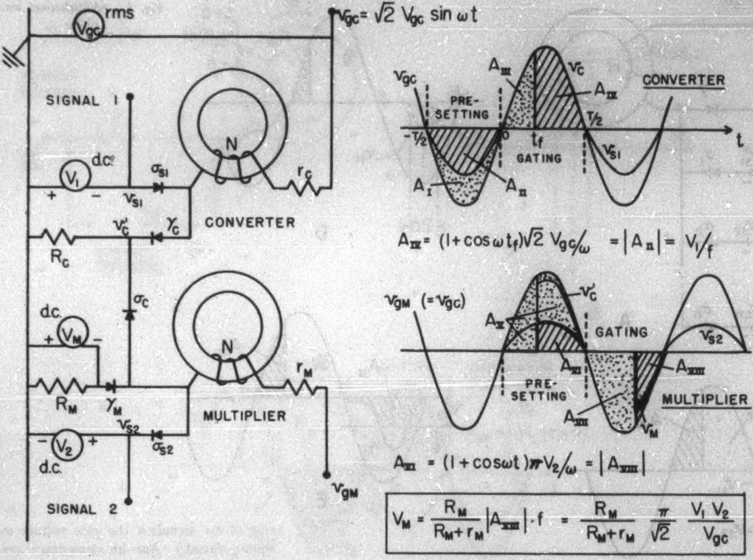


Fig. 3. Basic multiplying circuit

before, the wave forms of the signals being irrelevant as long as proper polarities insure blocking of the corresponding σ rectifiers. On the other hand, during presetting half-cycles the negative voltage v_g overrules whichever signal voltage happens to have the lowest absolute value, while γ and all the other σ rectifiers are blocking. That is, at any instant the rate of change of decaying core flux is dictated by the resultant of the gate voltage and of the weakest signal voltage. Accordingly, a magnetic coincidence counter[3] has been developed under the recognition that no output is found in the gating if even only one of the signal voltages has been zero through the presetting.

More generally it is seen here that the gated output area is equal to the area of the contour of least ordinates of gate and signal voltages during presetting as shown in Fig. 2(B). This general property may have a number of interesting analytical applications. If an assigned voltage $v_{s1}(t)$ is applied periodically in the negative half-cycles in conjunction with a large signal $v_{s2}(t)$ dropping to zero at $t=t_1$ as shown in Fig. 2(C), the reading of a D'Arsonval meter in the output is proportional to the integral

$$\int_0^{t_1} v_{s1}(t)\, dt$$

Thus if t_1 is gradually phase-shifted, a writing d-c meter of adequately large inertia can plot directly the indefinite integral function on a roll of paper moving synchronously with whatever organ controls the scanning phase shift of v_{s2}. (The periodical obliteration of the action of $v_{s1}(t)$ can be obtained and phase-shifted in various other ways, for example, by suitably timed mechanical contactors temporarily short-circuiting the rectifier γ in a single-channel amplifier).

Useful analytical information on the function $v_{1s}(t)$ as shown in Fig. 2(D) can be obtained also by use of a scanning wave v_{s2} of any chosen width $2a$ (or by contactors timely short-circuiting γ, for example at all times but $t_1-a \leq t \leq t_1+a$).

On the other hand, gross outputs equal to the areas shaded in Fig. 2(E) result from a scanning with progressively increased direct voltages v_{s2}; thus another type of information on the function v_{s1} is obtained. The simplicity of the circuits and components needed to perform these and similar analytical operations may be an incentive for actual usage in analogue computer studies, for example, as applied to statistical problems.

Multiplying Circuits

Two-Stage Multiplying Circuits

A single-channel amplifier of the type described in the preceding constitutes the converter stage shown in Fig. 3. This stage operates on a sinusoidal gate voltage $v_{gC}=\sqrt{2}V_{gC}\sin\omega t$ with a half-wave rectified signal voltage v_{s1}, of arbitrary wave form. This signal voltage is converted in the gross output voltage v_C absorbed by the resistances (R_C+r_C) in the subsequent half-cycle after firing.

A 2-channel amplifier of the type described in the preceding constitutes the multiplier stage. Its sinusoidal gate voltage is v_{gM}. (For the sake of simplicity v_{gM} is assumed identical to v_{gC} in Fig. 3.) A half-wave rectified sinusoidal signal voltage v_{s2} is one of the inputs, while the other input is the net converter output $v_C'=v_C R_C/(R_C+r_C)$ appearing across the load resistance R_C. In the time interval from 0 to t_f v_C' is zero (or nearly so) and therefore the presetting flux change of the multiplier core is dictated completely by v_{gM}, as the action of v_{s2} is obliterated over this interval. Therefore the total presetting over the half-cycle is limited only by the area A_{VI} of least ordinates, recognized with opposite sign as the area A_{VIII} of gross output of the multiplier in the next half-cycle. The various volt-time areas of the two stages are related easily to the readings of the rms voltmeter V_{gC} and of the d-c voltmeters V_1, V_2, and V_M as shown in Fig. 3, with the result

$$v_M = \frac{R_M}{R_M+r_M}\,\frac{\pi}{\sqrt{2}}\,\frac{V_1 V_2}{V_{gC}} \tag{1}$$

Thus multiplication as well as division is performed by the system.

It can be noted that V_{gC} appears in the denominator because increases of V_{gC} increase t_f and thus reduce the area A_{VI} and A_{VIII}. On the other hand, the crest of the multiplier gate voltage can be varied without any effect of first order on the results because such variations, while affecting the wave form of multiplier output voltage do not modify the volt-time area A_{VIII} and the reading of the meter V_M.

It may be noted also that the two stages do not need to have cores of equal dimensions or magnetic characteristics, and their windings can also have different turns, provided, of course, that in each stage the gate voltage satisfies the condition $V_{g\max}<\omega N\Phi_{\mathrm{sat}}$.

The operating range of the device has evident limits; namely $|v_{s1}|$ must be less than $|v_{gC}|$ during converter presetting, and the crest of v_{s2} must be lower than both $\sqrt{2}V_{gC}R_C/(R_C+r_C)$ and $\sqrt{2}V_{gM}$ during multiplier presetting. Different levels of signals can be handled by substituting any one input channel of the converter or of the multiplier with a separate signal winding of suitable turns $N_s \neq N$ in series with one half-rectifier. (The signal applied to any such winding then must act against the presetting action of a voltage $v_g'=v_gN_s/N$ impressed in series in the same winding from the secondary of a transformer with this turns ratio. Such known modifications of the elementary amplifier circuits of Figs. 1 and 2 do not modify in the least the basic operation of the multiplying system described here, and do not need to be examined in detail.)

Also it is recognized that the given signals v_{s1} and v_{s2} act upon the system in two subsequent half-cycles. Hence, a certain cross-correlation lag $\tau={}^1/_2f$ is inherent to the multiplication performed. (This lag may be increased, if so desired, for cross-correlation or autocorrelation studies, by the insertion of half-cyclic delay stages[2] between converter and multiplier.) One additional half-cycle lag affects the final output.

These considerations are not necessarily limited to sinusoidal gate voltage and signals. It is sufficient to point out that the circuit of Fig. 3 performs also multiplication of d-c signals if the gate voltages are square waves symmetrical about zero (as obtained from a d-c power source through a rotating commutator); in this case the various volt-time areas A_{I} to A_{VIII} become rectangles and the behaviors of converter and multiplier are even more clearly seen.

MULTISTAGE MULTIPLYING CIRCUITS

The output voltage V_M of the multiplier stage of Fig. 3 has a converted wave form of the same kind as v_c'. This suggests that a second multiplier stage can be added directly to the system of Fig. 3, one of its inputs being $v_M'=v_M$ while the other input is an assigned half-wave rectified sinusoidal signal voltage v_{s3}. The reasoning applied to the volt-time areas of the single multiplier stage of Fig. 3 can be extended to this new stage with the result

$$V_M''=\frac{R_M''}{R_M''+r_M''}\left(\frac{\pi}{\sqrt{2}}\right)^2\frac{V_1V_2V_3}{V_{gC}V_{gM}'} \qquad (2)$$

where

V_1, V_2, V_3 are d-c meter readings of the assigned signal voltages
V_M'' is the d-c meter reading of output voltage across the load resistance R_M'' of the second multiplier stage
V_{gC} and V_{gM}' are rms readings of the gate voltages of the converter and of the first multiplier.

This idea can be extended, in principle, to the use of any number μ of multiplier stages, each one of which receives the output voltage of the immediately preceding stage as a converted signal in conjunction with a new half-wave rectified sinusoidal signal independently assigned. The end output metered by $v_{M\mu}$ across the load resistance $R_{M\mu}$ of the last multiplier, is

$$V_{M\mu}=\frac{R_{M\mu}}{R_{M\mu}+r_{M\mu}}\left(\frac{\pi}{\sqrt{2}}\right)^{\mu}\times \frac{V_1V_2\ldots V_{(\mu+1)}}{V_{gC}V_{gM}'\ldots V_{gM(\mu-1)}} \qquad (3)$$

where the numerator factors are d-c meter readings of the assigned signals and the denominator factors are rms reading of gate voltages of all but the last stage. The various signals act upon the system in subsequent half-cycles, while one additional half-cycle lag affects the final output. Proper use of the concept of areas of least ordinates must be made to determine the upper limits of signal voltages allowable to control each stage effectively. The total number of stages practically usable in any such system may be limited, of course, by cumulative errors, some of which are discussed later.

EXPERIMENTAL RESULTS AND CRITICAL COMMENTS

Fig. 4 presents some experimental results of operations performed by the circuit of Fig. 3, slightly modified as sketched, with the following components:

1. Magnetic cores, Hypernik V; $DO=2{}^1/_2$ inches; $DI=1{}^1/_2$ inches; tape height $=1/2$ inch; number of turns per core $N=2{,}500$; winding resistances $r_C=r_M=65$ ohms.
2. Rectifiers: Germanium *G10*, 130 volts rms reverse voltage, 400 milliamperes (later use of selenium rectifiers yielded substantially identical results).
3. Load resistances $R_C=500$ ohms; $R_M=410$ ohms.

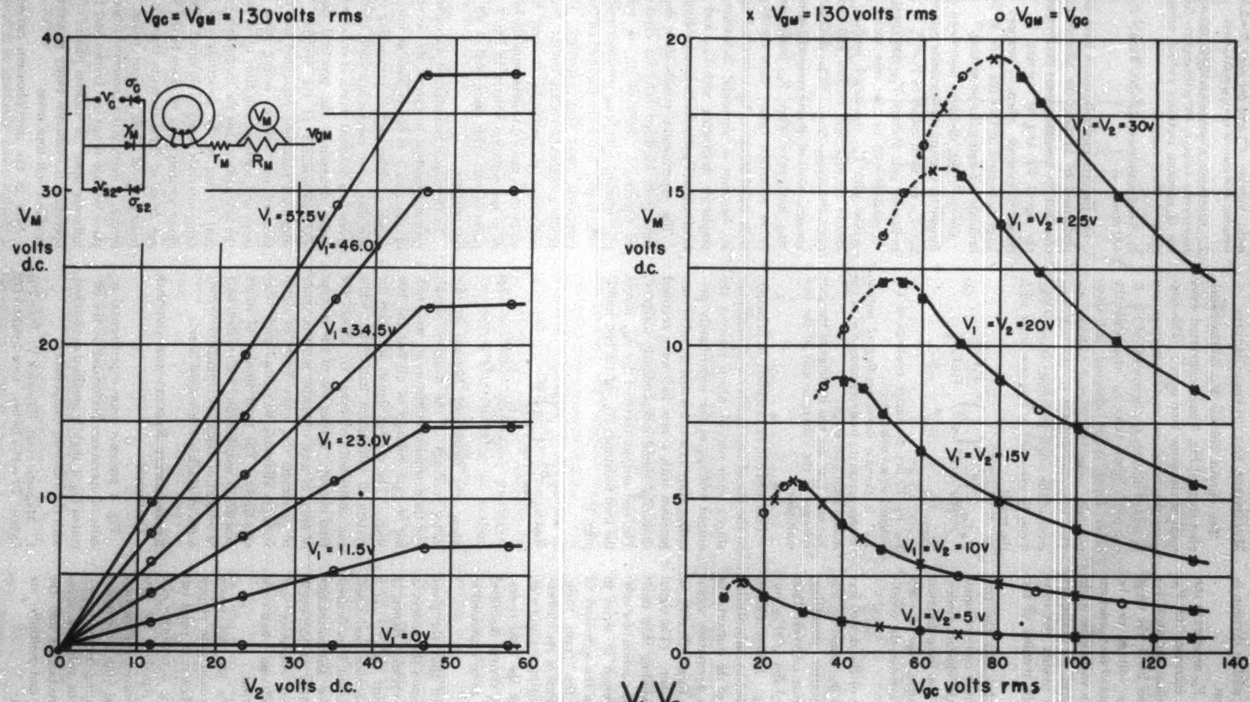
On the left multiplications have been performed by keeping the gate voltages constant, by assigning some fixed value to the signal V_1, and by varying V_2. V_M ceases to increase when V_2 becomes too large to control further the area of least ordinates of multiplier presetting.

On the right, divisions have been performed by varying the converter gate voltage V_{gC} for fixed values of $V_1=V_2$. The multiplier gate voltage V_{gM} was constant (crossed points) or else was varied together with V_{gC} (encircled points); as should be expected, crosses and circles fall on the same lines. (The rectifier σ_{s1} remains blocked during converter presetting half-cycles when V_{gC} is decreased below a certain value; then the hyperbolic trend breaks as the converted signal v'_C becomes a sinusoidal half-wave of amplitude decreasing with V_{gC}, and the final output V_M also decreases accordingly).

The experimental results show departures from idealized behavior expressed by equation 1. First of all, the multiplier winding after firing possesses residual inductance; hence the metered output is somewhat lower than expected. Practically a calibration of the circuit (simply performed, for example, at some large value of V_1 and V_2) to determine the factor K in the equation written in Fig. 4 disposes of this error quite well at all but low outputs.

In the experimentation it appeared of primary interest to extend the accuracy of equation 1 as far as possible into the region of either signal approaching zero. For this, more careful attention must be given to a different source of errors which really determines the practical design and performance of the system. The simplified reasoning presented has been focussed on voltages acting upon the cores and on their volt-time areas under the assumption that the magnetizing current flowing in the presetting and prefiring intervals is negligibly small. But a closer scrutiny must be given to the voltage drops of these currents through the resistances of their various paths. Of course, in every given numerical situation these voltage drops are known and their influence upon the various phases of the operation is evaluated easily in terms of modified volt-time areas.

Now in the multiplication curves of



$$V_M = K \frac{V_1 V_2}{V_{gc}}$$

Fig. 4. Some experimental results

Fig. 4 it is seen that for $V_1=0$ a certain output V_M, small but not quite zero, was metered. In fact, with $V_{s1}=0$, no firing of the converter occurs in its gating half-cycles, but still the flow of magnetizing current through R_C gives a certain converted signal v'_C to the multiplier; the latter's area of least ordinates A_{V1} is not brought to zero, and the multiplier does fire at some time near the end of its gating half-cycle with an output substantially independent of V_2. A direct use of the original circuit of Fig. 3 had caused an even larger error in the V_M reading because of the drop of prefiring magnetizing current through R_M. To avoid this, R_M was displaced in the position shown in the sketch of Fig. 4; in this manner its drop of magnetizing current over the range of low multiplier outputs nearly averages to zero in consecutive half-cycles. In particular, for any V_1, zero output is truly obtained when $V_2=0$. In the experimentation the rectified signal voltages were obtained from two variable transformers each feeding through a half-wave rectifier into a resistor R_{s1} (or R_{s2}) connected between ground and signal rectifier σ_{s1} (or σ_{s2}). (This is needed to provide a path for the presetting magnetizing currents flowing against the signal voltages.) In the experimental circuit R_{s1} and R_{s2} were chosen as $R_{s1}=R_{s2}=2{,}000$ ohms. As a consequence, at full output the power level of either signal source was about 1/4 of the gross output power. Since it is generally desirable to reduce the over-all signal power requirements, tests were also performed with $R_{s1}=R_{s2}=10{,}000$ ohms, with results not very different from those shown in Fig. 4, except for low values of V_2. (At low values of V_2 the wave form of v_{s2} was significantly distorted from the sinusoidal by the drop of magnetizing current.) Also with $R_{s1}=R_{s2}=2{,}000$ ohms, voltage measurements directly at the transformer terminals corresponded substantially to the readings of the meters V_1 and V_2 across the resistors. This was no longer true for $R_{s1}=R_{s2}=10{,}000$ ohms; in this case the experimental results plotted in terms of transformer voltages departed badly from linearity over the first one-third of the scale of the abscissae.

A more proper design of the circuit components would improve all this as the magnetizing currents (about ±1 milliampere) could be reduced by the use of cores of lesser mean length (the available cores had more than twice the diameter needed to accommodate the windings used.) The voltage drops in question could also be made less significant percentagewise by the use of higher gate and signal voltage levels, allowed by the core cross section but limited in the experimentation by rectifier ratings. And larger power outputs could have been obtained from a reduction of R_M. But the circuit described, while performing the operations assigned remarkably well, is not primarily a power amplifier.

Nevertheless a variety of means is available to circumvent this question. The operation from low-level signal sources can be made possible by means of input cathode followers. Also, if d-c signal voltages (and square-waved gate voltages) are available, no need is felt for the resistors R_{s1} and R_{s2} as the magnetizing current then can flow directly into the signal source.

Finally, one may aim more directly at a reduction of the magnetizing currents in question. Since the magnetomotive force required by a core over intervals of increasing or decreasing fluxes is practically constant, it may not be difficult to provide a large part of it by an auxiliary winding fed from some suitable current source during presetting and prefiring or, more simply, during presetting only. Procedures of this kind are mentioned in the references.

References

1. On the Mechanics of Magnetic Amplifier Operation, Robert A. Ramey. *AIEE Transactions*, vol. 70, 1951, pt. II, pp. 1214–23.

2. On the Control of Magnetic Amplifiers, R. A. Ramey. *Ibid.*, pp. 2124–28.

3. The Single-Core Magnetic Amplifier as a Computer Element, R. A. Ramey. *AIEE Transactions*, vol. 71, pt. I, 1952 (January 1953 section), pp. 442–46.

No Discussion

A Transient Analyzer for Magnetic Amplifiers

E. J. SMITH
MEMBER AIEE

Synopsis: The transient analyzer described is a simple and reliable electronic device which operates by generating a number of pulses exactly equal to the number of cycles required for the output current or voltage of a magnetic amplifier to change from its initial value to a predetermined value. The response time in cycles is obtained directly by recording the output of the analyzer with an electronic counter or conventional recording device.

THE RESPONSE TIME of a magnetic amplifier is defined as the time required for the rms or rectified average value of output current or voltage to change a prescribed percentage of the difference between the corresponding initial and final steady-state values. When the magnetic characteristic of the core material approaches the rectangular *B*—*H*-loop shape, the wave form of the output current exhibits the typical rapid rise as one core saturates at some angle α_{MA}, and sinusoidal form after saturation, until the end of the half cycle. Under these conditions, the peak value of load current is independent of the average or rms value when

$$\alpha_{MA} \leq \frac{\pi}{2}$$

Unfortunately, conventional recording instruments are better suited to the measurement of cyclic peak values, and can be adapted to the measurement of the rms or average value response time only at the expense of considerable inconvenience and difficulty. An electronic cyclic integrator which permits direct measurement of the rectified average value of the output

---

Paper 53-283, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 12, 1953; made available for printing April 23, 1953.

E. J. Smith is with the Polytechnic Institute of Brooklyn, Brooklyn, N. Y.

The study of Magnetic Amplifiers is being undertaken at the Microwave Research Institute of the Polytechnic Institute of Brooklyn, under the sponsorship of the Office of Naval Research contract N6ori-98, task order IV.

The assistance of Mr. Victor Boros is gratefully acknowledged.

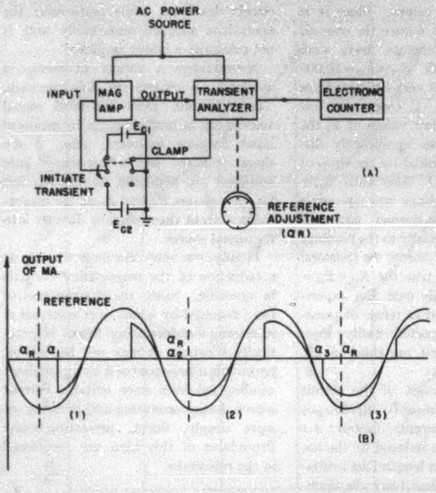

AC POWER SOURCE
INPUT
MAG AMP
OUTPUT
TRANSIENT ANALYZER
ELECTRONIC COUNTER
E_{c1}
CLAMP
INITIATE TRANSIENT
E_{c2}
REFERENCE ADJUSTMENT (α_R)
(A)
OUTPUT OF MA
REFERENCE
α_R
α_1
α_R
α_2
α_3
α_R
(1)
(2)
(3)
(B)

Fig. 1. Method of determination of transient response of a magnetic amplifier

A. Arrangement of magnetic amplifier and instruments for performing measurement

B. Typical output wave form of a magnetic amplifier during a transient

taken over each cycle has been previously described.[1,2] With this method, the response of the amplifier in terms of cyclic average values is portrayed on the screen of a cathode-ray oscilloscope, is photographed, and the response time determined in the usual manner.

The device to be described in this paper (called the transient analyzer) is also electronic, but is very simple and reliable as compared with the above device, and possesses accuracy of a high order. The output of the transient analyzer is a series of pulses equal in number to the number of cycles required for the load current to change from its initial value to any arbitrarily specified value, called the reference value. The reference may be specified in terms of rectified average or rms values by the use of a suitable meter, by performing a manual adjustment prior to the initiation of the transient.

Description of Transient Analyzer

PRINCIPLE OF OPERATION

The operation of the transient analyzer is best described by making reference to Fig. 1. The values of control-circuit voltage E_{c1}, E_{c2}, or current corresponding to the desired initial and final steady-state values of load current or voltage are first established. The output of the magnetic amplifier is then set at the reference value (that is, 63, 90, etc. per cent of the total change) by applying a suitable signal to the control circuit. Corresponding to the reference value of output, the magnetic amplifier saturates at some angle

$\alpha_{MA}=\alpha_R$

α_R is called the reference angle. When the core materials are not ideal the current wave does not jump abruptly at saturation; however, if α_{MA} is taken as the angle of maximum slope of the wave, the operation is the same as though the materials were rectangular.

The transient analyzer generates a voltage pulse of very short duration once every cycle. The phase of this reference pulse is next made to coincide with the reference angle by a manual adjustment. When the output current is less than the reference value

$\alpha_{MA}>\alpha_R$

and when the output current is greater than the reference value

$\alpha_{MA}<\alpha_R$

(Although these remarks apply only to a single-ended magnetic amplifier, the method itself can be applied to a push-pull amplifier by comparing the saturation angle of one reactor with the reference angle.)

The function of the transient analyzer is to compare the angles α_{MA} and α_R every cycle, and to indicate the result by generating an output pulse every cycle if $\alpha_{MA}>\alpha_R$, but no pulse if $\alpha_{MA}<\alpha_R$ for a build-up transient; and by generating a pulse every cycle if $\alpha_{MA}<\alpha_R$, but no pulse if $\alpha_{MA}>\alpha_R$ for a decay transient. The transient analyzer becomes operative only after the switch initiating the transient is thrown. Therefore, the number of pulses generated by the analyzer after the switch is thrown is equal to the number of cycles required for the magnetic amplifier output to change from the initial value to the reference value. The response time of the magnetic amplifier is obtained directly by counting the number of pulses generated by the analyzer during the transient. An electronic counter of the conventional type (multiple decade counters employing Eccles-Jordan flip-flop circuits, as used in electronic digital computers) was found to be a very convenient means for counting the pulses; however, a standard-type recording oscillograph could be used equally as well, but with perhaps less convenience.

The comparison of the angles α_{MA} and α_R is made during one half of the cycle

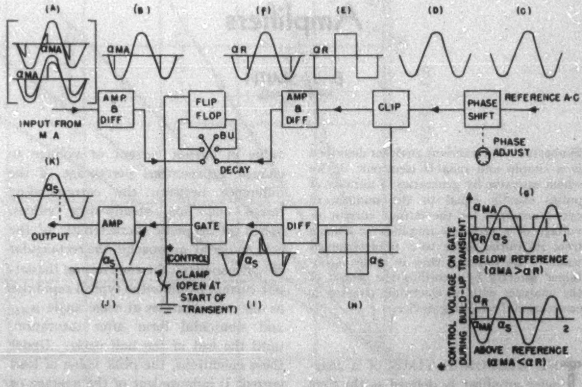

(A)
(B)
(F)
(E)
(D)
(C)
αMA
αMA
αMA
αR
αR
AMP. & DIFF
INPUT FROM M A
FLIP FLOP
BU
DECAY
AMP & DIFF
CLIP
PHASE SHIFT
REFERENCE A-C
PHASE ADJUST
(K)
αS
OUTPUT
AMP
GATE
DIFF
CONTROL
CLAMP (OPEN AT START OF TRANSIENT)
(J)
(I)
(H)
αS
αS
αS
(G)
αMA
αR
αS
1
BELOW REFERENCE (αMA > αR)
αR
αMA
αS
2
ABOVE REFERENCE (αMA < αR)
CONTROL VOLTAGE ON GATE DURING BUILD-UP TRANSIENT

Fig. 2. Block diagram of the transient analyzer, illustrating wave forms at various points

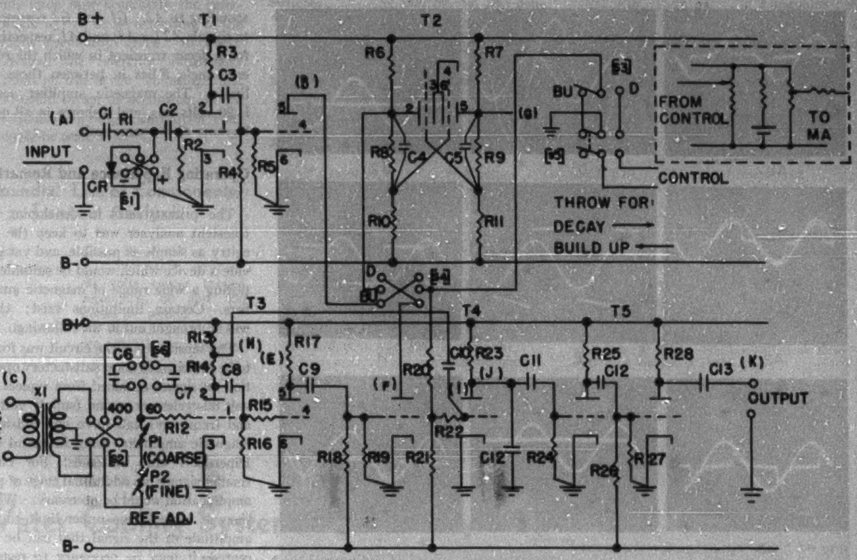


Fig. 3. Circuit diagram of transient analyzer

(the negative half in Fig. 1). If the output of the magnetic amplifier is alternating current, as shown in Fig. 1, the signal into the transient analyzer can be obtained from the load (that is, load voltage). If the signal is full-wave direct current (rectified direct-current), then the signal to the transient analyzer must be taken as the voltage across one reactor winding (see waves in upper left corner of Fig. 2). In any case (that is, full-wave direct current, full-wave alternating current, half-wave, or push-pull alternating current or direct current), the signal can always be taken from a winding of one reactor. The reason for the above is that the analyzer operates by comparing the reference pulse with some negative pulse derived from the magnetic amplifier which occurs only once during a cycle. This requirement is always met by taking the signal from a winding on any core in the amplifier. The input impedance of the transient analyzer is high enough (one megohm) to prevent detrimental loading of the circuits.

The reference pulse phase angle α_R is readily adjusted without the aid of an oscilloscope as follows:

1. The output of the magnetic amplifier is set at the reference value by applying an appropriate steady-state control signal, and observing the output with a suitable a-c or d-c meter.
2. The transient analyzer is unclamped and the phase adjustment varied until pulses are obtained from the output terminals. (The electronic counter counts continuously as indicated by neon bulbs.)
3. The phase adjustment is then varied (coarse and fine adjustments) until the output from the analyzer just ceases. The reference angle α_R is now adjusted properly.
4. The analyzer is again clamped, the counter reset to zero, and the transient initiated by throwing the switch. At the completion of the transient, the reading of the counter is exactly equal to the response time in cycles.

DESCRIPTION OF CIRCUITS

The internal operation of the transient analyzer can best be explained by reference to the block diagram of Fig. 2. The input signal *A* is clipped,* differentiated, and amplified, resulting in a sharp negative pulse of short duration *B*, which occurs at phase angle α_{MA}. An a-c voltage *C* having the same frequency as the magnetic supply is passed through a phase shifter, which is equipped with coarse and fine adjustments for varying the phase. The phase shifted a-c wave *D* is clipped *E*, differentiated, and amplified, resulting in a sharp negative pulse of short duration *F;* this is the reference pulse. The phase of the reference angle α_R is determined by the phase adjustments previously described. The two pulses *B* and *F* are fed into opposite plates of a conventional Eccles-Jordan flip-flop circuit. The operation of the Eccles-Jordan[3] circuit is such that a negative pulse received at one plate triggers the stage, causing that plate voltage to drop to its low value and the voltage of the opposite plate to rise to its high value. The circuit then remains in this condition until a negative pulse is received at the plate of the second (opposite) tube, at which time the plate voltage values of the two tubes reverse.

The voltage from one plate of the flip-flop circuit *G* is used to control the gate stage. When this plate is high the gate is open and allows to pass through it a signal appearing at its input terminals *I*. When this plate is low the gate is closed and the incoming signal *I* is blocked.

The two input connections to the flip-flop circuit are interchanged for build-up and decay transients. For a build-up transient, the pulse derived from the magnetic amplifier *B* is connected to trigger the flip-flop in such a direction as to open the gate, and the reference pulse

---

* The negative half cycle is removed if the negative drop in the signal occurs on the positive half cycle (input from a reactor winding) and the positive half cycle is removed if the negative drop occurs during the negative half cycle. See (*A*) in Fig. 2; for circuit details see Fig. 3.

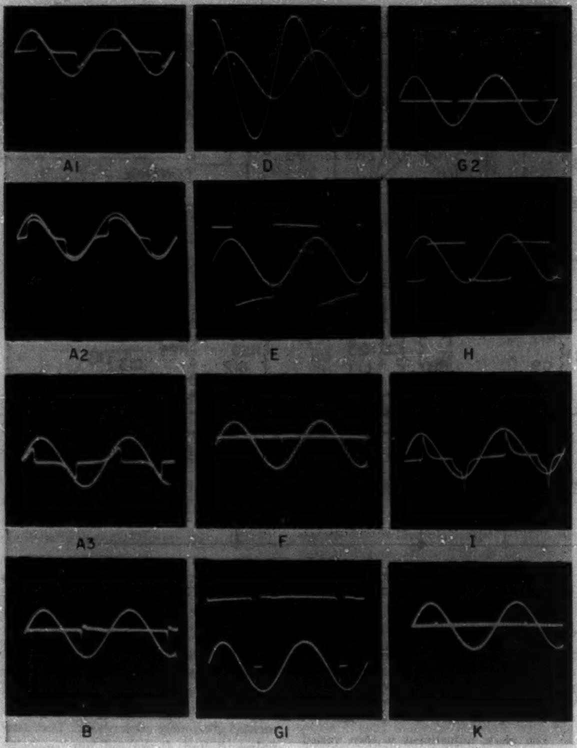


Fig. 4. Oscillograms of voltage wave forms at the various points in the circuit and block diagrams

F is connected in such a direction as to close the gate. The signal input (called the sensing pulse) to the gate *I* occurs at angle α_s, where

$$\alpha_s = \alpha_B + \pi$$

It follows immediately that if *B* occurs after *F*, that is

$$\alpha_{MA} > \alpha_B$$

the gate will be open when the sensing pulse *I* arrives at the input to the gate; therefore, a pulse appears at the output *J* and *K* of the device. On the other hand, during a build-up transient, if *F* occurs after *B*, that is

$$\alpha_{MA} < \alpha_B$$

the gate will be closed when the sensing pulse *I* arrives at the input to the gate; under these conditions, no pulse appears at the output of the transient analyzer. For a decay transient, the roles of the *B* and *F* pulses are interchanged by throwing the reversing switch shown in Fig. 2. The operation of the transient analyzer then proceeds as previously described.

The complete circuit diagram of the transient analyzer is shown in Fig. 3. Conventional pulse-circuit techniques are used; therefore, no detailed description is necessary. The particular circuit design shown is intended primarily for 60 and 400-cycles-per-second operation, but could be modified for higher frequency use with minor changes. Wave forms of voltages at various points in the circuit are shown in Fig. 4. (Oscillograms *B* and *F* are obtained with the tube T_3 removed from its socket.) The letters *A* to *K* refer to the circuit points shown in Figs. 2 and 3. *A1* and *A2* are the output voltages across the load of a Hipernik-V full-wave a-c amplifier for two values of control signal; *A3* is the voltage across the load winding of one reactor corresponding to *A2*. *G1* and *G2* correspond to signals *A2* (or *A3*) and *A1*, respectively for a decay transient in which the reference angle *F* lies in between these two limits. The magnetic amplifier source is 20 volts rms, and is shown in all oscillograms to the same scale.

Operating Experience and Remarks

The primary aim in developing the transient analyzer was to keep the circuitry as simple as possible, and yet provide a device which would be suitable for testing a wide range of magnetic amplifiers. Certain limitations exist; these will be brought out in the following.

The sensitivity of the circuit was found to be sufficient to allow satisfactory operation on signals obtained from typical 10-volt 60-cycles-per-second (supply voltage and frequency) half-wave and full-wave magnetic amplifiers, using Mumetal and Hipernik-V core materials. For much smaller signals, an additional stage of preamplification would be necessary. While there is virtually no upper limit to the amplitude of the signal that can be accepted, it may be necessary to restrict the amplitude in certain cases, where distortions or spurious pulses exist in the output of the magnetic amplifier wave; this can always be done with a high-resistance voltage-divider network.

The logic of the circuit operation requires that the input from the magnetic amplifier be sufficient to trigger the flip-flop stage for all values of α_{MA} realized during the transient. This condition becomes difficult to fulfill when either the initial or final value of α_{MA} approaches π, even though the signal strength may be adequate over the major portion of the transient. While the logic of the design could be changed to take care of this situation, the additional circuit complexity makes the change undesirable and contrary to the original aim of simplicity. Furthermore, the case for which α_{MA} approaches π is a very unusual one, and its exclusion does not impose a serious limitation on the usefullness of the device. At this point it may be asked whether the aforementioned limitation on α_{MA} does not exclude the application of the transient analyzer to push-pull magnetic amplifiers of the null-balance type, in which the initial or final steady-state output is usually zero. The answer is negative, because with such circuits the signal is obtained not from the output, but from a winding on one core. Hence, even at null balance, the signal α_{MA} is not likely to approach π.

The accuracy of the transient analyzer

depends upon how accurately the reference angle α_R can be set according to the procedure previously given. In every case tested, the reference adjustment (the point at which the "counting" just ceases) could be set as accurately as the meter (indicating output current or voltage) could be read.

Appendix. List of Components—Transient Analyzer

Tubes	*T1* to *T5*—*6SN7*
Transformer	*X1*—120 volts primary, 80 volts center tapped secondary, 60 and 400 cycles per second
Potentiometers	*P1*—850 kilohms, 4 watts *P2*—20 kilohms, 4 watts
Crystal Diode	*CR*—*IN34*
Switches	*S1*, *S2*, *S4*, *S5*, *S6*—double-pole-double-throw, no neutral position *S3*—single-pole-double-throw, neutral position
Power Supply	*B*+, 150 volts, 40 milliamperes *B*−, 100 volts, 10 milliamperes
Resistors	(1 watt, carbon, ±5 per cent)
R13, *R14*, *R28*	10 kilohms
R3, *R17*, *R25*	18 kilohms
R6, *R7*	24 kilohms (matched within 2 per cent)
R5, *R19*, *R27*	100 kilohms
R23	150 kilohms
R8, *R9* and *R10*, *R11*	240 kilohms (matched within 2 per cent)
R2, *R15*, *R21*, *R22*	470 kilohms
R1, *R18*, *R20*, *R24*, *R26*	1 megohm
R4	1.3 megohms
R12, *R16*	2.2 megohms
Condensers	(±10 per cent)
C11	56 micromicrofarads
C4, *C5*	100 micromicrofarads
C9	230 micromicrofarads
C12	360 micromicrofarads
C3, *C10*	0.001 microfarad
C2	0.002 microfarad
C8	0.003 microfarad
C6	0.007 microfarad
C7	0.05 microfarad
C1, *C13*	0.1 microfarad

References

1. A CYCLIC INTEGRATOR, S. Roman. Thesis, Polytechnic Institute of Brooklyn, N. Y., 1951.

2. A STUDY OF MAGNETIC AMPLIFIERS. PART IV—TRANSIENT RESPONSE, E. J. Smith. Report R-238 50, PIB-183, Microwave Research Institute, Polytechnic Institute of Brooklyn, N. Y.

3. THEORY AND APPLICATIONS OF ELECTRON TUBES, H. J. Reich. McGraw-Hill Book Company, New York, N. Y., 1944.

No Discussion

Water Cooling Systems of Mercury-Arc Rectifiers

AIEE COMMITTEE REPORT

Synopsis: This report is intended as a guide on the design, application, and operation of water cooling systems for mercury-arc rectifiers. Its contents are based on the accumulated experience acquired by rectifier manufacturers and users over many years. The report deals with the various engineering and operating aspects of rectifier cooling systems. Several types of water cooling systems are described, and their application is discussed. The various factors influencing corrosion are reviewed. The characteristics of cooling waters are discussed and recommendations are given on the chemical treatment of coolants, the application of antifreeze solutions, and the maintenance requirements of cooling systems. The results of a survey on the operation and maintenance of cooling systems in representative rectifier installations are also included.

THERE are two general types of mercury-arc rectifiers in operation in the United States and Canada: the multianode type, with 6, 12, or 18 anodes with a common mercury cathode in one tank; and the single-anode type, with one anode and its cathode in an individual tank. Nearly all new rectifier installations made in the last 10 years or longer are of the single-anode ignitron or excitron type. For this reason, the report refers primarily to the single-anode type of rectifier; however, its contents apply generally to both types.

A standard rectifier consists of an assembly of 6 or 12 tubes with the auxiliary equipment. The rectifier cooling system, which is the subject of this report, is one of the essential auxiliaries, and its performance is an important factor in the operation of the rectifier.

The flow of current through a mercury-arc rectifier tube is accompanied by a voltage drop called the arc drop, which may have a value in the range of 15 to 25 volts, depending on the tube design and rating. The product of the current and arc-drop voltage is a power loss which is converted into heat. Most of the heat is transferred to the cooled surfaces of the tube by radiation and by condensation of mercury vapor.

For satisfactory operation, the heat generated by the losses has to be removed and the tube temperature maintained in the optimum operating range. The temperature of the cooled surfaces controls the mercury vapor density in the tube. Excessive temperatures are likely to cause arc backs. If the temperature is too low, the arc-drop voltage may increase and voltage surges might be produced due to arc starvation, when the mercury vapor density is insufficient for conduction of the tube current.

Glass-bulb rectifier tubes and some metal-tank tubes are cooled by air forced over the tube surfaces, with a fan or blower. However, most of the rectifiers in the United States and Canada are cooled by circulation of water through water jackets or tubular coils. Some tube designs have copper tubing soldered on the outside surface; others have internal steel coils. Combinations of external and internal coils or of internal coils and external jackets are also used. The water-cooled parts of sealed tubes are usually made of stainless steel.

Rectifier Cooling Systems

The following four types of water cooling systems are used for mercury-arc rectifiers: direct raw water cooling system; direct raw water cooling system with recirculation; heat exchanger cooling

---

Paper 53-288, recommended by the AIEE Electronic Power Converters Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 17, 1953; made available for printing April 27, 1953.

The personnel of the Working Group on Water Cooling Systems of Mercury Arc Rectifiers of the AIEE Committee on Electronic Power Converters are: E. J. Remsheid, *Chairman*, Mrs. B. O. Buckland, A. P. Colaiaco, A. G. Dickinson, G. N. Hughes, J. Klotz (alternate), J. B. Nayler, R. J. Nix, M. W. Rew, Mark Sareault, and Harold Winograd.

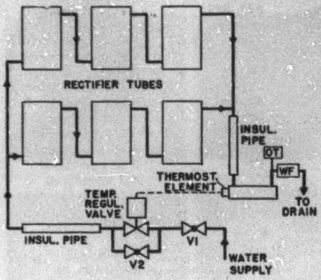


Fig. 1 (left). Direct raw-water cooling system

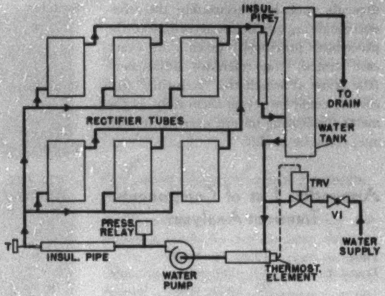


Fig. 2 (right). Direct raw-water cooling system with recirculation

system, using a water-to-water heat exchanger; and heat exchanger cooling system, using a water-to-air heat exchanger.

These cooling systems are illustrated in Figs. 1 through 4. The figures show typical cooling circuits of a 6-tube rectifier, without regard to the type of cooling elements on the tubes or the physical location of the component parts of the system. The cooling circuits are applicable also to rectifiers with a smaller or larger number of tubes. Individual designs of cooling systems may differ from the illustrations in various details.

Direct Raw Water Cooling System

This cooling system is illustrated in Fig. 1. Raw water from an available water supply flows through the rectifier tubes and is discharged to the drain. The flow of water is regulated by an automatic temperature regulating valve, which is actuated by a thermostatic element on the discharge side. The regulating valve conserves water and maintains the tube temperatures in the desired range.

The regulating valve is shunted by the small hand valve *V2* to permit the flow of a minimum amount of water when the regulating valve is closed. This minimum flow is needed for actuating the regulating valve by the thermostatic element, when the water discharge temperature increases and additional water is required. The cooling system is connected to the water supply through hand valve *V1*.

The rectifier tubes are at the potential of the cathode terminal, which is connected to the positive side of the d-c load circuit supplied by the rectifier. With rare exceptions, the positive side of the d-c circuit is insulated from ground. On some applications the negative side is grounded; on others, there is a floating ground connection between the positive and negative poles. For this reason the water connections between the rectifier and the grounded water supply and drain are made through insulating pipes, such as rubber hose or plastic tubing. The insulating connections are made sufficiently long to limit the flow of leakage current from the water to ground to a few milliamperes in order to minimize the electrolysis of the water and the corrosive effect of the liberated oxygen.

For traction systems the negative side of the d-c circuit is usually grounded. This practice has been recommended by the Technical Practices Committee of the National Association of Corrosion Engineers.[1]

To obtain good heat transfer between the cooled tube surfaces and the water, a specified minimum rate of water flow is required for each tube at the rated loads and overloads. To provide the required flow for this type of cooling system without using an exesssive amount of water the tubes are usually connected in series or in a series parallel combination as shown in Fig. 1.

For protection of the tube, an overtemperature relay *OT* on the discharge side and a water flow relay *WF* or a pressure relay are usually provided to give an alarm signal or take the rectifier out of service if the water flow is insufficient or has stopped completely. In automatic stations an electrically operated water valve is used for shutting off water when the rectifier is not in service.

Direct Raw Water Cooling System with Recirculation

This type of cooling system is shown in Fig. 2. The cooling water is recirculated between the rectifier tubes and a water tank by a motor-driven water pump at a constant rate. Raw water is admitted to the cooling circuit by a temperature regulating valve *TRV* as required for maintaining the water temperature in the desired range. The surplus water is discharged to the drain through an overflow pipe on the water tank. When the water temperature is below the setting of the thermostatic element, the regulating valve is closed.

The regulating valve may be connected to admit the raw water to the water tank instead of the piping shown in Fig. 2. The overflow opening in the tank should preferably be located above the highest point of the recirculating system, in order to keep the water passages of the system full and prevent admission of air when the pump is not running.

The rectifier tubes are usually connected in parallel. The water pump should have sufficient capacity to supply the required flow rate of the tubes. The tubes are connected to the grounded part of the cooling circuit through insulating pipes. As an alternative, the insulating pipes may be placed between the regulating valve and the hand valve *V1* on the water supply side, and between the water tank and the drain. The equipment of the cooling circuit would then be at the rectifier potential.

A pressure relay connected to the inlet side of the tubes provides protection against operation of the rectifier without water circulation. The contacts of the pressure relay may be connected in the rectifier control system to prevent starting the rectifier unless water is circulating through the tubes, and to shut it down if the water flow should stop accidentally. A thermometer *T* is provided for indicating the water temperature.

Heat Exchanger Cooling System with Water-to-Water Heat Exchanger

In this cooling system, illustrated in Fig. 3, the rectifier cooling water is circulated between the tubes and the soft-water passages of a water-to-water heat exchanger. The recirculated water is cooled in the heat exchanger by the flow of water from a supply through the raw-water passages. The heat transfer takes place through the walls separating the soft-water and raw-water passages. The heat exchanger usually consists of a copper-alloy tube bundle enclosed in a cast-iron or steel shell. The recirculated water flows through the shell across the tube bundle, while the raw water flows through the tube bundle.

In Fig. 3, the rectifier water circuit is indicated by heavy lines and the raw water circuit by lighter lines. The rectifier tubes are connected in parallel. The water is circulated between the tubes and the heat exchanger at a constant rate by a motor-driven water pump. The flow of raw water through the heat exchanger is controlled by a temperature regulating valve *TRV* in the water supply circuit so as to maintain the rectifier water temperature in the desired operating range. The regulating valve is actuated by a thermostatic element in the rectifier cooling circuit. A pressure relay on the inlet side of the tubes provides protection against operation without water circulation. Thermostats are usually installed on the outlet side of the tubes for overtemperature protection; either a common thermostat or an individual thermostat for each tube is used. A thermometer *T* on the inlet or outlet side of the tubes indicates the water temperature.

A water expansion tank is connected to the rectifier cooling system on the intake side of the pump and is mounted above the highest point of the water circuit. Its purpose is to keep the cooling system full and prevent entrance of air; it serves as a reservoir to permit expansion of the water when hot and to make up for loss of water due to leaks. The expansion tank is provided with an opening to the atmosphere; an overflow pipe may be connected to this opening. The water level in the tank is indicated by a level gauge.

If water from the raw-water supply is used in the recirculating system, a hand valve *V2* may be provided for filling the system or for adding water when necessary.

On some rectifier designs the heat exchanger and other components of the cooling system are mounted on the rectifier assembly at the rectifier potential, as indicated by the dot-and-dash rectangle in Fig. 3. The connections to the grounded water supply and drain are then made through insulating pipes. If the rectifier water is treated with a corrosion inhibitor, it is preferable to maintain the cooling equipment at the rectifier potential, even if this equipment is mounted away from the rectifier, to reduce the leakage current to ground, because the treated water has a much lower resistivity than the raw water. In some installations, it is necessary to ground the cooling equipment to facilitate maintenance and to ensure safety.

Where it is necessary to ground the cooling equipment, the insulating pipes are inserted in the water circuit between the grounded equipment and the rectifier. In such a case untreated distilled water, which has a high resistivity, may be used in the recirculating system, or the insulating connections may be made sufficiently long to keep the leakage current within reasonable limits if the water is treated.

Heat Exchanger Cooling System with Water-to-Air Heat Exchanger

This type of cooling system, illustrated in Fig. 4, is similar to the cooling system of Fig. 3, except that a water-to-air instead of a water-to-water heat exchanger is used. The heat exchanger consists of parallel tubes connected between headers and provided with fins on their outside surfaces. The water passing through the tubes is cooled by the forced flow of air over the tubes, supplied by a motor-driven fan or blower.

One method of controlling the water temperature is to regulate the flow of water through the heat exchanger by a double-valve temperature regulator *TRV* as shown in Fig. 4. One valve *S* is in series with the heat exchanger; the other valve *P* is in parallel with it. The thermostatic element controls the positioning of the valves so that one closes while the other opens, which regulates the proportion of the rectifier water passing through the heat exchanger. The operation of the fan may be controlled automatically by limit switches on the valve operator or by high-low thermostat coordinated with the setting of the temperature regulator.

Another method of controlling the water temperature is to start and stop the fan by high-low thermostats. In some installations several heat exchangers with individual fans are connected in parallel to the water circuit, and the tem-

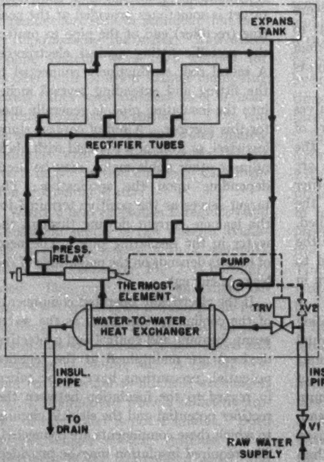


Fig. 3 (left). Heat exchanger cooling system with water-to-water heat exchanger

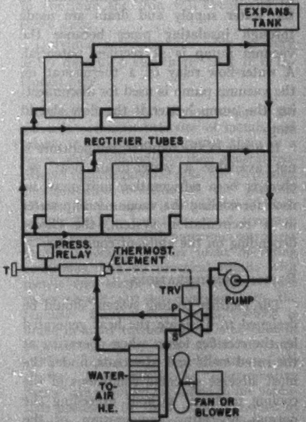


Fig. 4 (right). Heat exchanger cooling system with water-to-air heat exchanger

perature is controlled by the sequential starting and stopping of fans. Still another method of controlling the water temperature is to regulate the flow of air by controlling the speed of the fan motor or the position of air dampers, but this method is rarely used.

To limit the leakage current from ground, it is preferable to insulate the entire cooling system from ground, particularly if the water is treated with a corrosion inhibitor or an antifreeze solution. If the heat exchanger is located away from the rectifier, as is usually done, insulating pipes should be used in the connections of the rectifier cooling system. This provides protection against the flow of power current in case of accidental grounding of the heat exchanger; it also permits temporary grounding for inspection and maintenance. The heat exchanger may also be permanently grounded for safety reasons. In some installations the heat exchanger is placed outdoors.

COOLING OF VACUUM PUMP

The mercury-condensation vacuum pump used with pumped rectifiers and the water-cooled vacuum pipes require a small quantity of cooling water, approximately 1 gallon per minute or less. The vacuum pump requires water of lower temperature than the rectifier, and its cooling circuit is independent of the rectifier cooling system. For efficient operation of the vacuum pump, the cooling water temperature should be below 35 degrees centigrade. Raw water is generally used.

The cooling circuits of the vacuum pump and vacuum pipes are usually connected in series. The connections to the water supply and drain are made through insulating pipes because the vacuum pump is at rectifier potential. A water-flow relay or a thermostat on the vacuum pump is used for disconnecting the pump heater if the flow should stop.

If water of the required temperature is not available a water-to-air heat exchanger or a refrigeration unit may be used for cooling the vacuum pump water in a recirculation system, the choice depending on the ambient conditions.

DESIGN CONSIDERATIONS

The rectifier cooling system should be designed to dissipate the heat generated by the rectifier losses when operating at the rated loads and overloads, under the most adverse specified conditions of the cooling medium, without exceeding the normal operating temperature of the rectifier. The operating temperature, measured on the water discharge side of the tubes, is generally in the range of 45 to 55 degrees centigrade, depending on the tube design and the direct voltage.

The water-to-water heat exchanger should be selected for the highest specified temperature and the lowest specified pressure of the raw water supply. The water-to-air heat exchanger should be selected for the highest specified cooling air temperature. If the water-to-air heat exchanger is placed indoors, provision should be made in the building for the intake and discharge of the required quantity of cooling air. The intake and discharge openings should be far enough apart so that the discharged air will not re-enter through the intake. If ducts are used for the air, the blower should have sufficient capacity to take care of the pressure drop in the ducts in addition to the pressure drop of the heat exchanger.

The water pump should be selected to circulate the required rate of water flow against the total friction head of the cooling circuit. The flow rate is not critical; however, if it is too low the heat transfer in the heat exchanger as well as in the rectifier tubes is reduced. The motor driving a centrifugal water pump is usually matched to the pump so that the motor cannot be overloaded under any flow condition from blocked flow to open discharge. In designing the water circuit, care should be taken to have a positive pressure (above atmospheric) at the pump intake in order to prevent entrance of air through the shaft seal of the pump. To satisfy this condition, the static head from the point of atmospheric pressure (the expansion tank of Figs. 3 and 4) to the pump intake should be higher than the friction drop between these points of the circuit.

In recirculating cooling systems, valves should be provided at the lower part of the circuit for filling and draining the system. Small valves or pet cocks are usually provided on the pump and other parts of the cooling circuit to vent the air trapped when the system is filled. Air in the system can block or reduce the flow of water. Water may be added to the system either by a valve, as shown in Fig. 3 or at the expansion tank.

Valves may be provided in the cooling circuit to permit servicing individual items, such as the water pump, without draining the system. The number of such valves should be kept to a minimum because they increase the friction head. Gate valves should preferably be used because they have less friction drop than globe valves. The latter type is used where the water flow has to be throttled, as in the raw-water supply circuit.

For the cooling systems of Figs. 1 through 3, the size of the temperature regulating valve should be selected to suit the flow requirements at the prevailing water pressures. It should be large enough to permit the required water flow, when fully open, at the lowest operating pressure. However, if it is oversized and permits the required water flow when barely open, it will "hunt" and will not provide satisfactory temperature control. If the pressure of the raw-water supply varies over a wide range, a pressure regulator may be required. Strainers are provided in installations where the raw water contains filterable solids.

For some rectifier applications, water heaters are installed in the cooling circuit to maintain the water temperature within the desired operating range. The control of the heaters should be interlocked with the flow or pressure relay, or the pump motor, so that they can be energized only when there is a flow of water.

When a rectifier is on load, some of the internal parts of the tubes, such as the anodes, attain fairly high temperatures and continue to transmit heat to the cooling water for some time after the rectifier is shut down. For this reason the cooling system should continue to operate for 30 minutes or longer after the rectifier is shut down.

Where an insulating pipe is employed in the cooling system between points having a d-c potential difference, as shown in Figs. 1 through 3, an electrolytic target is sometimes provided at the positive (rectifier) end of the pipe to protect the metallic fitting against electrolysis. A metal rod, conductively connected to the fitting and extending several inches into the insulating pipe, is generally used for this purpose. A rod of copper, stainless steel, or iron, or a rod clad with platinum or other noble metal, may be used, depending upon the application. The target serves as the positive terminal for the leakage current flowing through the water in the insulating pipe. The need of targets depends on the magnitude of the leakage current.

If the electrically operated components of the cooling system (such as the water pump, heaters, and control and protective devices) are maintained at the rectifier potential, precautions have to be taken in regard to the insulation between the rectifier potential and the electric circuits to which these components are connected. The required insulation may be provided within the individual devices or through insulating transformers, depending on the operating voltage.

In a 6-phase double-way rectifier circuit, three tubes have a common cathode connection, while the other three tubes have a common anode connection and their cathodes are at different potentials from all the other cathodes. For this reason each of the latter tubes has to be connected to the cooling system through individual insulation pipes.

APPLICATION

The direct raw-water cooling system, Fig. 1, is the simplest and least expensive of the four types described. Because of the minimum water flow required at all times, the operating temperature of the rectifier tubes is likely to be below the optimum value at light loads, and moisture may condense on the outside tube surfaces when the humidity is high. For applications with a low load factor, the water consumption may be higher than for the other cooling systems, due to the minimum water flow requirement.

The raw-water cooling system with recirculation, Fig. 2, is second in regard to cost. It is less expensive than the heat-exchanger systems because the heat exchanger is omitted; the other components of the system are substantially the same. This is the most economical system in regard to water consumption, and the tube temperatures can be maintained in the optimum range.

Neither raw-water cooling system permits the use of a corrosion inhibitor or an antifreeze solution. The water passages of the tubes have to be cleaned periodically to remove any scale or sediment that may be deposited by the water. Raw-water cooling systems have been applied for rectifier units of lower ratings employing sealed tubes with stainless steel or nonferrous water-cooled surfaces. They can be used only when the raw-water quality conforms with the purity specifications given in the section entitled "Classification of Cooling Water."

Where a heat exchanger is required and a sufficient supply of raw water is available, a cooling system with a water-to-water heat exchanger, Fig. 3, is normally used. This type of heat exchanger occupies less space than a water-to-air heat exchanger, requires no fan, and no provision has to be made for the intake and discharge of cooling air. A cooling system with a water-to-air heat exchanger, Fig. 4, is employed when a sufficient supply of raw water is not available. Both heat-exchanger cooling systems permit treatment of the recirculated water, and the deposition of scale or sediment in the closed system is minimized.

Corrosion

The corrosion of metals may be defined as their destruction by the chemical action of their environment and must not be confused with erosion which is destruction by mechanical means. The rusting of iron is the most common example of corrosion, and sand blasting of iron is an example of erosion.

In the cooling systems normally employed for rectifiers, erosion is usually a negligible factor, whereas corrosion is of prime importance and requires diligent consideration of both design and maintenance of the units to insure a satisfactory service life. The types of corrosion commonly encountered in rectifier cooling systems are: atmospheric, galvanic, electrolysis, and acid attack.

ATMOSPHERIC CORROSION

Atmospheric corrosion or, more commonly, the reduction of iron to iron oxide, will form a thin film of oxide on the iron when exposed to dry air at ordinary temperatures. This film effectively isolates the iron from the oxygen and greatly reduces the rate of oxidation. However, if this film is removed in any way, the oxidation will continue.

Iron in an aqueous solution tends to go into solution as iron ions. However, in doing so, it must displace an equivalent number of hydrogen ions which gather on the surface of the iron as a thin invisible film. This film of hydrogen retards the tendency of the iron to go into solution and thus slows up the reaction. However if this film of hydrogen is removed, either by combining with oxygen or by forming hydrogen bubbles, the reaction will continue, the iron going into solution being thrown down as rust.

In an acid solution, the tendency of the hydrogen to plate out on the iron increases with the acidity and may increase the gas pressure in the hydrogen film sufficiently to form bubbles. This will destroy the protective hydrogen film and permit the iron to continue to go into solution. Thus, corrosion of iron is in general much more rapid in acid solutions than in alkaline solutions.

In a water cooling system which contains a continuous supply of oxygen, iron will combine with the oxygen to form hydrated ferric oxide which will precipitate as brown rust, and corrosion may be quite serious. This can occur in a direct raw-water cooling system, in which the oxygen dissolved in the water is liberated in the water passages. Copper, alloys of copper such as brass, also aluminum, and stainless steel form protective oxide coatings and are much more resistant to corrosion.

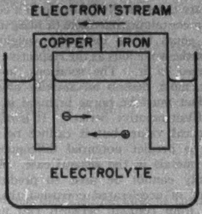


Fig. 5. Galvanic corrosion

In closed or recirculating cooling systems where the supply of oxygen is limited, the oxygen is gradually removed by combining with the iron, or by liberation and escape through vents, and the rate of corrosion is retarded. Also, in a closed system the solution can be made alkaline preventing the destruction of the protective hydrogen film on the iron and further retarding corrosion.

GALVANIC CORROSION

When two dissimilar metals, such as copper and iron, are brought into electric contact and are connected by an electrolyte, a current will flow between them, Fig. 5. The two metals in contact constitute in effect a miniature battery or galvanic cell with the iron as the anode. In the solution, electrons will flow from the copper to the iron and iron ions will flow from the iron to the copper. It is apparent that if this condition is allowed to continue the iron will eventually be eaten away.

Galvanic corrosion has been so very aptly described elsewhere[3] that it is included here.

Galvanic Series—With a knowledge of the galvanic corrosion behavior of metals and alloys, it is possible to arrange them in a series which will indicate their general tendencies to form galvanic cells, and to predict the probable direction of the galvanic effects. Such a series is provided in Table I.

This Table is based on actual corrosion testing experience with numerous corrosives in the laboratory, under plant operating conditions, and on practical results with metals and alloys in long service. This series should not be confused with the familiar "Electromotive Series," which is found in many textbooks and is of value in physical chemistry and thermo-dynamic studies. Let us make this difference a little clearer. The metals in the "Electromotive Series" are arranged according to the potentials possessed by them when immersed in solutions of a definite strength. These solutions are theoretically made up in a concentration containing one mole (a mass numerically equal to its molecular weight) per liter of the same ions as the respective metal immersed in it. The solutions, in other words, must possess an activity of unity. This fact must be borne in mind in using the "Electromotive Series." It is of great theoretical value, but it cannot be relied upon to predict potential differences between metals in the general case. And it certainly cannot be used to predict the amount of accelerated corrosion that will result from contact between two metals. Table I has basic characteristics which make it analogous to the "Electromotive Series," but it also takes into consideration over-all and practical aspects in addition to theoretical principles. The potentials measured in these practical corrosive environments do determine which one of two metals will be anodic and which cathodic.

Table I. Galvanic Series of Metals and Alloys

Corroded End (anodic, or least noble)

Magnesium
Magnesium alloys
Zinc
Aluminum 2S
Cadmium
Aluminum 17ST
Steel or Iron
Cast Iron
Chromium-Iron (active)
Ni-Resist
18–8 Stainless (active)
18–8–3 Stainless (active)
Lead-Tin solders
Lead
Tin
Nickel (active)
Inconel (active)
Brasses
Copper
Bronzes
Copper-Nickel alloys
Monel
Silver solder
Nickel (passive)
Inconel (passive)
Chromium-Iron (passive)
18–8 Stainless (passive)
18–8–3 Stainless (passive)
Silver
Graphite
Gold
Platinum

Protected End (cathodic, or most noble)

It will be noticed that some of the metals in Table I are grouped together. These group members have no strong tendency to produce galvanic corrosion on each other, and from the practical standpoint they are relatively safe to use in contact with each other. But the coupling of two metals from *different* groups and *distant* from each other in the list will result in galvanic, or accelerated, corrosion of the one higher in the list. The farther apart the metals stand, the greater will be the galvanic tendency. This may be determined by measurement of the electrical potential difference between them, and this is often done. But it is not practical to tabulate these differences, because the voltage values for combinations of the metals will vary with every different corrosive condition. What actually determines galvanic effect is the quantity of current generated, rather than the potential difference.

The relative position of a metal within a group sometimes changes with external conditions, but it is only rarely that changes occur from group to group. It will be seen that the chromium-iron and chromium-nickel-iron alloys are in two places in the Table. They frequently change positions as indicated, depending upon the corrosive media. The most important reasons for this are the oxidizing power and acidity of the solutions, and the presence of activating ions, such as halides. Inconel and nickel also occasionally behave in a similar manner, though the variations of their position are less frequent and less extensive. In environments where these alloys ordinarily demonstrate good resistance to corrosion, they will be in their passive condition and behave accordingly in galvanic couples.

Factors Influencing Galvanic Corrosion—The galvanic series indicates only the tendency of the several metals and alloys to set up galvanic corrosion. Actual corrosion cannot proceed unless there is a flow of electrical current. It is basically a question of amperes rather than volts. Just as in ordinary corrosion, the ease with which this current can flow and the magnitude of the current are controlled by a number of basic factors.

These factors are generally discussed under the headings, (1) Conductivity of the Circuit, (2) Potential Between the Anode and the Cathode, (3) Polarization, (4) Relative Cathode and Anode Areas, (5) Geometrical and Relationship Between Dissimilar Metal Surfaces, and (6) Contact Between the Metals.

A practical demonstration of corrosion by galvanic action occurs at the junction of an iron and a copper pipe in a cooling system. As shown in Fig. 6, the iron, or the less noble metal of the two, forms the anode of a galvanic couple and is corroded, with the deepest penetration at the contact point of the two metals.

Aluminum should not be used in a cooling system with copper or its alloys because of their wide separation in the galvanic series. Aluminum, being closer to the anodic end, would be subjected to corrosion.

There are various other causes of corrosion by galvanic currents and these have been listed[2] as follows:

1. Contact with metals.
2. Presence of impurities.
3. Grain boundaries, which are frequently anodic towards grain interiors.
4. Differential thermal treatment, important near welds.
5. Local scratches and cut edges, which are often anodic to the rest of the surface.
6. Differential stress or strain.
7. Differential concentration of salts, especially in the case of copper.
8. Differential aeration.
9. Differential heating.
10. Differential illumination.

Most of these items are not as important to mercury-arc rectifier cooling systems as contact between dissimilar metals. However, differential aeration is one which is thought by some to have a serious influence.

An example of galvanic corrosion by differential aeration is shown in Fig. 7. A small cavity or crack exists in the surface of the metal. Oxygen cannot diffuse readily to the bottom of this crack, whereas there is an abundance of it at the surface of the metal. A current is produced between the aerated surface of the metal (cathodic) and the unaerated crevice which is anodic. Since the attack is concentrated in the small crevice, and assuming that there is an abundant supply of oxygen at the surface, the rate of corrosion will be very great, usually becoming greater as the depth of the pit is increased. Thus, in designing cooling systems it is important to avoid such recesses.

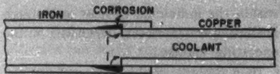


Fig. 6. Galvanic corrosion of water pipe

Corrosion by Electrolysis

In many rectifier installations the negative side of the d-c circuit is grounded or there is a ground connection at some intermediate point between the positive and negative terminals, so that the rectifier tubes operate at a positive voltage with respect to ground. If the coolant supply is grounded, it is brought to the rectifier through an insulating pipe and direct current flows through the coolant to ground. The magnitude of this current is determined by the impressed voltage, the diameter and length of the insulating hose, and the resistivity of the cooling medium.

The metal at the (anodic) positive end of the insulating pipe will be subject to electrolytic attack and may be eaten away to the extent that eventually a leak will result. Another effect of this current is to generate oxygen at the anode and hydrogen at the cathode. The oxygen so generated will tend to aggravate corrosion in other parts of the system.

One method used to prevent damaging electrolytic attack at the anodic end is to insert an electrolytic target which projects into the insulating section of the piping. These targets carry the major portion of the electrolytic current and suffer most of the deterioration, the small leakage current which is carried by the pipe fitting doing negligible damage.

Fig. 8 shows such a target arrangement. The target can be replaced when it becomes so short that it no longer provides adequate protection for the fittings. The rate at which the target wears away is proportional to the electrolytic current. In accordance with Faraday's law, one gram of equivalent (the weight of a sub-

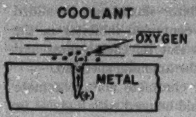
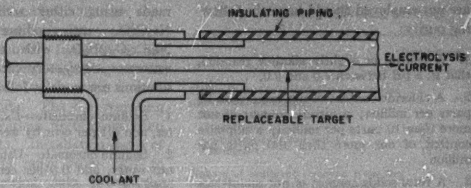


Fig. 7 (left). Galvanic corrosion by differential aeration

Fig. 8 (right). Electrolytic target



stance displacing or otherwise reacting with 1.008 grams of hydrogen or combining with 1/2 gram atomic weight—8.00 grams—of oxygen) weight of metal will be removed for each 96,500 coulombs (1 coulomb is a current of 1 ampere flowing for 1 second). Thus, 27.9 grams of iron will be corroded for each 96,500 coulombs or approximately 20 pounds per ampere-year. The choice of target materials depends on the degree of protection required, and the coolant being used. Iron, stainless steel, platinum, and copper have been used successfully.

In cases where an inhibitor is used in the cooling system for protection against atmospheric corrosion, the inhibitor usually greatly decreases the resistivity of the coolant with a proportional increase in electrolytic current and wearing away of the target. One commonly used inhibitor is sodium dichromate and borax added to the water in the ratio of 0.5 per cent by weight each. This reduces the resistivity of the cooling water to a small fraction of that of distilled water. However, the protection afforded the rest of the system by the inhibitor more than justifies the additional wear on the targets.

Corrosion by Acid Attack

Most common metals corrode much more rapidly in an acid solution than they will in a neutral or slightly alkaline solution. When a metal is placed in an acid solution it will displace hydrogen ions from the solution and its own ions will go into solution. The noble metals will displace hydrogen ions very slowly, forming a thin film of hydrogen ions on the surface of the metal, which will prevent further dissolving of the metallic ions into the solution, and corrosion is stopped. Metals such as gold, platinum, silver, and copper fall into this category.

The base metals such as iron, zinc, aluminum, and magnesium liberate hydrogen in an acid solution much more rapidly, to the extent that the hydrogen bubbles off in a stream rather than forming a protective film on the metal. This process will continue with the metallic ions continually going into solution, and the parent metal will gradually dissolve away. Metals such as lead, tin, and nickel liberate hydrogen very slowly in acid solutions and corrode very slowly.

A solution which is neither acid nor alkaline, therefore neutral, will have a pH value of 7. A cooling solution with a pH value between 7 and 9 (slightly alkaline) is satisfactory for the cooling systems of mercury-arc rectifiers. It is also important that the cooling medium should not be allowed to become excessively alkaline because metals such as aluminum and zinc, whose oxides are soluble in alkalies, will corrode rapidly in a highly alkaline solution.

Classification of Cooling Water

By the application of filtration, distillation, and chemical treatment, any kind of water can be prepared for a specific purpose. In order to classify raw waters for coolant purposes, factors such as color, pH, turbidity, and chemical analyses should be known. It is not necessary that raw waters, used as rectifier coolants, have a potable bacteriological count. Potable waters are usually free of many troublesome contaminants, but the question of using them as a coolant in many localities is an economic one.

Hardness in waters is caused almost entirely by the presence of calcium and magnesium. To a lesser degree, hardness is increased by minerals such as iron, aluminum, strontium, barium, zinc, and free acids. Hard waters are less corrosive than soft waters, but are more conducive to scaling. In terms of $CaCO_3$ in parts per million very hard waters have a hardness of 201 to 500; moderately hard, 101 to 200; slightly hard, 56 to 100; and soft waters are termed as 0 to 55 parts per million.

Raw-water coolants for power rectifiers are classified as surface waters, ground waters, and purified waters.

Surface Waters

Surface waters are taken from rivers, ponds, lakes, or ocean bays. The quantity of contaminants in such waters is sometimes high, requiring straining or filtration to some degree. At times chlorination is necessary to prevent the growth of bacteria and algae in the finely divided suspended organic matter. Waters pumped from ocean bays are considered brackish because of their high salt content. Desirable low-temperature surface waters are obtained when the intake pipe is at a considerable depth.

Surface waters are applied most frequently as the raw water in cooling systems with water-to-water heat exchangers. Relatively clean surface waters are used in direct raw-water cooling of rectifiers. A high quality potable surface water is obtained from the Great Lakes for many of our larger cities and is well suited for either direct or heat-exchanger cooling or for filling recirculating systems.

Ground Waters

Ground waters are those pumped from the earth, as from artesian wells. Well waters differ greatly in hardness and chemical composition, the differences being due to the different kinds of mineral matter the water contacts while percolating and flowing through the ground.

Like surface waters, ground waters of suitable quality are well adapted to water-to-water types of heat exchangers and for direct raw-water cooling. Ground waters have a desirable characteristic in that the intake temperature is several degrees below the particular ambient.

When applying ground water as a coolant, it is important that the chemical analysis be studied and the possible degree of scaling noted. Chemical analyses are not always available to assist in an appraisal of cooling water. In such cases, an electrical resistivity measurement of the water will provide a satisfactory guide to the total amounts of dissolved solids. Waters having a resistivity of 2,500 ohm-centimeters or higher, when measured at about 25 degrees centigrade, are usually satisfactory as coolants. When desirable, the approximate amount of total dissolved solids in parts per million can be determined by the equation

$$\text{Total dissolved solids in parts per million} = \frac{640{,}000}{\text{specific resistivity in ohm-centimeters}}$$

The recommended quality of cooling water as specified in the American Standards Association Standard *C-34.1-1949*[4] is as follows: Water for direct raw-water cooling systems, where heat exchangers are not employed should have the following purity:

1. A neutral or slightly alkaline reaction, that is, a pH between 7.0 and 9.0.
2. A chloride content of not more than 20 parts per million; a nitrate content of not more than 10 parts per million; a sulphate content of not more than 100 parts per million.
3. A total solids content of not more than 250 parts per million.
4. A total hardness, as calcium carbonate, of not more than 250 parts per million.

PURIFIED WATERS

Distilled water, steam condensate, and properly deionized waters are usually referred to as purified waters. Purified waters are applied specifically in heat-exchanger cooling systems where a coolant of high electrical resistivity is required, as in cases where the water flows through insulating pipes between points of high potential difference.

Waters having a resistivity greater than 30,000 ohm-centimeters are considered purified waters. Steam condensate and deionized water usually have a resistivity of 30,000 ohm-centimeters or higher. Distilled water has a resistivity of 100,000 ohm-centimeters or higher. The resistivity of purified water in a rectifier cooling system is gradually lowered while in use, and the water may have to be replaced periodically if a certain minimum resistance is required.

Chemical Treatment of Rectifier Coolants

Chemical treatment of water in recirculating cooling systems is resorted to for two reasons: 1. To prevent corrosion of metal parts. 2. To prevent freezing of the coolant.

CORROSION INHIBITOR

Various treatments have been tried, but a solution of sodium chromate and water has been found to be the most economical and practical method of preventing the corrosion of the iron parts of a rectifier cooling system. Corrosion is inhibited by the deposition on the metallic surfaces of a dense and adherent protective film of iron oxide and chromic oxide, Cr_2O_3, produced by a simple corrosion reaction. The protective film forms quickly, and the local deposition current gradually decreases as the thickness of the film and its resistance increases. One desirable feature of the protective film is that it is self-healing and will retard any action tending to cause corrosion.

Corrosion inhibiting solutions can be made using either sodium chromate, Na_2CrO_4, or sodium dichromate, $Na_2Cr_2O_7 \cdot 2H_2O$, with either distilled waters or satisfactory tap waters. Such solutions are made as follows:

1. Sodium chromate—Using distilled water, add 0.1 per cent by weight of Na_2CrO_4.
2. Sodium chromate—Using a satisfactory tap water, add 0.5 per cent by weight of Na_2CrO_4.
3. Sodium dichromate—Using distilled water, add 0.5 per cent by weight of $Na_2Cr_2O_7 \cdot 2H_2O$ and 0.5 per cent by weight of borax. Since sodium dichromate tends to increase the acidity of water, the borax is added to increase the alkalinity.
4. Sodium dichromate—Using a satisfactory tap water, add 1 per cent of $Na_2Cr_2O_7 \cdot 2H_2O$ and 1 per cent of borax.

These chromate solutions have proved satisfactory and are recommended in the American Standards Association Standard *C-34.1-1949*.[5]

ANTIFREEZE

Antifreeze solutions are used in recirculating rectifier cooling systems which are exposed to freezing temperatures. This applies most frequently to cooling systems with water-to-air heat exchangers.

There are two general types of antifreeze materials, those with a glycol base and those with an alcohol base. The latter are divided into two classes, methanol base and ethanol base. The glycol base antifreezes have a higher boiling point and are therefore referred to as permanent type antifreezes. Since rectifiers operate at relatively low temperatures, either type of antifreeze is satisfactory. The methanol base types are less expensive and smaller amounts are required for the same temperature protection.

Most commercial antifreezes contain corrosion inhibitors which normally give adequate protection for approximately 3 months. At this time the system should be checked, and if the solution is rusty the inhibitor has lost is effectiveness.

The presence of alcohol in the system without an inhibitor will loosen the protective film of the metallic surfaces. If an alcohol base antifreeze is continued in use after its corrosion inhibitor has lost its effectiveness, an inhibitor should be added. Borax can be used as an inhibitor for an ethanol base antifreeze; a 0.5 per cent solution by weight of borax (0.7 ounce per gallon of solution) is adequate. For a methanol base antifreeze, either borax or sodium chromate may be used as an inhibitor.

The presence of too much inhibitor in the system may prove detrimental and therefore inhibitor should not be added until the self-contained inhibitor in the antifreeze has lost its effectiveness and then only in the amounts specified.

Neither the chromate nor dichromate inhibitors should be used with a glycol or ethanol alcohol base antifreeze because serious corrosion will result.

A commercial synthetic methanol of 99.85 per cent purity (which does not contain any inhibitor) may be used as an antifreeze in a rectifier cooling system in which sodium chromate is used as an inhibitor. The percentage of methanol is determined by the lowest temperature for which protection is required.

Operation and Maintenance

The amount of maintenance a cooling system will require depends much upon the nature of the particular coolant. The more important maintenance items and operating precautions are outlined in the following.

It is necessary to check periodically, at intervals of 6 months or less, the effectiveness of the corrosion inhibitor. This can be done either by checking the electrical resistivity of the coolant or by checking its actual corrosion retarding qualities in an electrolytic cell. The electrolytic cell is provided with two iron wire electrodes with approximately 5 volts d-c potential applied across them. With the electrodes partly immersed in the liquid, electrolytic disintegration of the positive electrode will be noted in a few minutes should the coolant be of a weak concentration. When a sufficient amount of sodium chromate is present, no disintegration will be noted even if the test is continued for many hours. The frequency of the tests will depend upon the amount of make-up water added to the recirculating system, which would naturally weaken the solution.

When the coolant is checked by the resistivity method, it is suggested that all tests be made at a specific temperature. The following values indicate the permissible range when tests are made at a temperature of 25 degrees centigrade:

0.1 per cent solution of sodium chromate, 600 to 800 ohm-centimeters

0.5 per cent solution of sodium chromate, 150 to 200 ohm-centimeters

Because of the nature of a chromate solution, minute water leaks will not seal themselves. It is therefore necessary to make all pipe joints watertight prior to the use of the chromate solution. The cooling system should be inspected frequently for water leaks.

Systems utilizing raw water, such as sealed tube rectifiers or the raw-water side of a heat exchanger, must be cleaned periodically of scale or water sediment. Iron parts of a raw-water system will corrode and in time must be replaced. The cleaning intervals may vary from 3 months to 2 years, depending on the quality of the water.

Excessive amounts of scaling not only will reduce the water flow but will impair the heat-transfer efficiency of cooling surfaces. The removal of scale from a cooling system can be done safely by any of several commercial chemical methods, such as by recirculating solutions of Rodine or oxalic acid.

Cooling tubes of a water-to-water heat exchanger, provided with removable headers, can be cleaned mechanically with wire brushes or metal rods.

It is obvious that all accessory equipment of the cooling system, such as thermal devices, pressure relays, sight gauges and automatic water valves must be maintained in satisfactory working condition. Most of these devices will be conditioned when a cooling system is chemically cleaned. Devices comprising moving parts, such as water pumps, automatic water valves, and fan motors, require special attention.

When a rectifier is in operation, each tube must receive sufficient water to maintain its temperature within the specified operating limits. If the flow of water through an individual tube is too low because of a partial obstruction, it would be indicated by a higher temperature on its water discharge side as compared to the discharge temperature of the other tubes. Air trapped in the cooling circuit may obstruct the flow of water.

For satisfactory protection against freezing, it is recommended that antifreeze solutions be discarded and replaced once each year. Frequent tests of specific gravity should be made to insure adequate protection during the winter months. If electrolytic targets are used, they should be checked periodically and replaced when they become so short that there is danger of electrolysis at the pipe fittings.

All cooling systems have a number of controls to regulate the tube temperature and to act as safety devices. These should be checked once a year for operation and settings.

Survey of Rectifier Cooling Systems

QUESTIONNAIRE RETURNS

A questionnaire on rectifier cooling systems was sent to 81 rectifier users. Returns were received from 32 users, representing 475 operating ignitron, excitron and multianode units. Included in the number are rectifiers operating in all classes of service, both in the United States and Canada. The d-c operating voltages of the various units range from 250 to 1,800 volts. Installation dates of the various units range from 1927 to 1951. The following data reveal the number of rectifier units involved for each specific question asked.

TYPES OF COOLING SYSTEMS

The three usual methods of cooling rectifiers are by raw water, water-to-water heat exchangers, or water-to-air heat exchangers. The number of rectifier units in the survey employing each of the three types of cooling systems is as follows:

Rectifiers cooled by water-to-water heat exchangers, 291

Rectifiers cooled by water-to-air heat exchangers, 174

Rectifiers cooled by raw water (sealed tube units), 10

TREATMENT OF RAW WATER

Chemical treatment of raw water depends specifically on the quality of water and local conditions. So-called hard waters are treated to remove scale-forming materials. Waters having a comparatively low pH value are treated to make them less corrosive. Fungus growth is prevented by chlorine treatment. The number of rectifier units using chemically treated raw water is 86.

Water containing suspended matter requires filtration. The survey indicates that in some cases this is done either by the use of filters or settling tanks.

TREATMENT OF RECIRCULATING COOLANT

Chemical treatment of recirculating coolants is used specifically to retard or prevent the corrosion of iron parts of the cooling system. The survey indicates the following interesting facts regarding the recirculating coolants used in 475 operating units. The percentages of the units employing the various coolants are:

	Percentage
Raw water, untreated	29.0
Distilled water	27.2
Steam condensate	10
Sodium chromate and water	31.5
Other chemicals and water	2.3

CLEANING OF COOLING SYSTEMS

To maintain the effectiveness of the heat-transfer surfaces of heat exchangers and rectifier tubes, periodic cleaning is necessary. Rectifier manufacturers have designed the various components of coolant systems to facilitate this maintenance procedure. Because of the difference types of coolants and the specific differences in cooling systems, the cleaning methods employed and the type of sediment removed vary with the installations. This variance is portrayed by the percentages given in Table I, indicating the preferences of various rectifier users.

Table I. Variance in Cleaning Methods

	Percentage
Methods used in cleaning the raw-water side of heat exchangers:	
Wire brush	28.5
Compressed air	26.0
Compressed air and water	34.0
Chemical	11.5
Nature of sediment removed from the raw-water side of heat exchangers:	
Hard scale	17.5
Loose deposit	65.0
Tubercles	17.5
Methods used in cleaning the recirculating side of heat exchangers:	
Flushing with water	41.0
Compressed air and water	38.5
Compressed air	12
Chemical	8.5
Nature of sediment removed from the recirculating side of heat exchangers:	
Hard scale	18.3
Loose deposit	59.0
Tubercles	22.7
Methods used in cleaning water tubes in the water-to-air heat exchangers:	
Wire brush	14.2
Compressed air	85.8
Nature of sediment removed from the water tubes of water-to-air heat exchangers:	
Loose deposit	
Methods of cleaning water jackets or coolant piping of rectifier tubes:	
Wire brush	29.0
Sand blast	8.5
Scraper	20.5
Compressed air and water	29.8
Chemical	12.2
Nature of sediment removed from water jackets or coolant piping of rectifier tubes:	
Hard scale	32.2
Loose deposit	48.5
Tubercles	19.3

It is noted from the percentages given in Table I that raw waters containing mineral matter which form loose deposits are often used as rectifier coolants. Fortunately such deposits can be removed most readily by reverse flushing or with compressed air and water. Hard scale is met with in some instances and requires harsher removal methods, such as a wire brush, sand blast, scraper, or chemical means.

When hard scale is encountered, water softening by the removal of calcium and magnesium salts is an economical remedy. The bulk of the calcium salts is usually removed by treatment with lime and sodium carbonate, and the magnesium salts are treated with sodium hydroxide.

Tubercles or large pieces of metal oxides often grow in parts of a cooling system. Water flow loosens these masses, which are then carried through the system until they bind themselves in narrow passages and thus impair the performance of the system. Tubercles of this nature usually form on iron surfaces when untreated water is used. Many tubercles grow to a size 10 to 30 times the amount of metal corroded in producing them. The addition of an alkali to the coolant usually prevents the growth of tubercles.

ELECTROLYTIC TARGETS

The survey indicates that 348 of the 475 rectifier units reported use some form of electrolytic target. The target materials include copper, steel, zinc, copper zinc alloy, iron, carbon, and platinum. The usual location of such targets is at the rectifier ends of insulating water pipes.

PAINTING OF VACUUM CHAMBERS

Many multianode rectifier users find it desirable to paint the water-cooled tank surfaces periodically. The types of coating applied are Apexior paint, graphite base paint, and aluminum powder mixed in varnish. Such painting, when applied to clean metal surfaces, will give protection from corrosion for approximately 2 years.

GROUNDING OF HEAT EXCHANGERS

Approximately 55 per cent of the survey returns indicate a preference for operating the heat exchangers at rectifier potential. This arrangement has several advantages. No insulating water pipes are required between the component parts; also the heat exchanger and rectifier can be mounted on a common base. Thirty-eight per cent of the returns indicate a preference for keeping the heat exchangers at ground potential. One advantage of this arrangement is that some maintenance work can be done on the heat exchanger while the rectifier is energized.

MAINTENANCE

The following items were reported as requiring more than the usual amount of maintenance:

1. Regulating water valves, both motor-operated and bellows-operated.
2. Thermostats.
3. Early deterioration of insulating water pipes, especially rubber hoses.
4. Cooling jackets on mercury-vapor vacuum pumps.
5. Bearings of pumps and fans.

SUMMARY

A study of the questionnaire returns indicates that materials in suspension and materials which are classed as dissolved solids in raw water involve more maintenance hours than do actual corrosion troubles. Such materials cause erosion of pump impellers and piping, deposition of scale and mud in the system, and deposition of a heat-insulating film on the heat-transfer surfaces of heat exchangers and rectifier tubes. Such materials also interfere with the functioning of control devices, cloud gauge glasses, and sight flow indicators. Seasonable changes in the chemical nature of raw waters are also a cause for additional maintenance hours.

It is gratifying to note, however, that cooling systems provided with sodium chromate operate for many years free from corrosion troubles.

Appendix I. Heat Transfer and Fluid Flow

A rectifier functions as an electronic switching device, for it neither generates nor stores energy. The heat liberated comes entirely from the arc loss during conduction between anode and cathode. The thermal capacity in a rectifier tube is low, and the cooling system must be proportioned to dissipate all losses readily whether generated in continuous or intermittent duty.

The required operating temperature and rate of water flow is specified by the tube manufacturer, and the cooling equipment supplied with the rectifier is designed to maintain the prescribed operating temperature. There is an optimum temperature range for satisfactory operation, and the cooling system is designed to keep the temperature within that range. Unlike some other types of electric equipment, the rating of a rectifier is not determined by the temperature rise of the water-cooled parts, and the loading cannot be increased by lowering the temperature below the specified values. Operating at excessively low temperatures can be detrimental to the rectifier.

Scale formation interferes with the water flow, as it restricts the flow area. It also interferes with heat transfer, as it introduces an additional thermal resistance between the water-cooled surfaces and the water. This makes the wall temperature hotter for the same water flow and temperature. The heat-transfer coefficients for scale deposits are given by McAdams.[6]

The greater part of the heat liberated at the anode of a rectifier tube is dissipated by radiation to the cooled surface of the tube. Part of the heat is dissipated by conduction to a radiator or bus terminal connection, which maintains the anode seal at a safe operating temperature. A large part of the heat generated at the cathode is dissipated by condensation of the vaporized mercury at the cooled surface.

Water Cooling

In a water cooling system the water flow required to dissipate the losses can be calculated from the following equation

$$\text{gpm}=\frac{3.8\ \text{kw}}{t_{\text{outlet}}-t_{\text{inlet}}}$$

where

gpm = gallons per minute

$$\text{kw}=\frac{\text{arc loss (volts)}\times\text{load current (amperes)}}{1{,}000}$$

t = temperature, degrees centigrade

In some cases, as in a direct raw-water cooling system, the available head of water is one of the system limitations, and thus determines whether the flow paths are to be connected in series or in parallel.

Air Cooling

In an air cooling system the air flow required to dissipate the losses can be calculated from the following equation

$$\text{cfm}=\frac{1{,}750\times\text{kw}}{t_{\text{outlet}}-t_{\text{inlet}}}$$

where cfm = cubic feet of air per minute at sea level.

For air cooling the air velocity must be much higher than the corresponding velocity for water cooling. Consequently, an air system requires a substantial head measured in inches of water.

Conversion Factors

1 kw = 3,413 Btu per hour = 56.88 Btu per minute = 239.1 calories per second
1 Btu = 0.000293 kilowatt-hour = 0.293 watt-hour
1 gpm water = 500 pound-hours
1 cfm air = 4.5 pound-hours
Specific heat of water = 1.0 Btu per pound (degrees Fahrenheit) or calories per gram (degrees centigrade) at normal cooling temperatures
Specific heat of air = 0.24 Btu per pound (degrees Fahrenheit) or calories per gram (degrees centigrade) at normal cooling temperatures
1 pound per inch^2 = 0.434 foot head of water

References

1. REPORT ON ELECTRICAL GROUNDING PRACTICES. *Report A-TP12*, National Association of Corrosion Engineers, Houston, Tex.

2. ASPECTS OF GALVANIC CORROSION. *Inco Corrosion Reporter*, International Nickel Company, Inc., New York, N. Y., vol. 2, no. 3, 1950.

3. GALVANIC CORROSION, R. B. Mears, R. H. Brown. *Industrial and Engineering Chemistry*, Easton, Pa., vol. 33, 1941, p. 1001.

4. POOL CATHODE MERCURY-ARC POWER RECTIFIERS. *ASA C-34.1—1949*, American Standards Association, New York, N. Y., section 4.400, 1949.

5. *Ibid.*, section 4.402.

6. HEAT TRANSMISSION (book), W. H. McAdams. McGraw-Hill Book Company, New York, N. Y., 1942, chap. VI, p. 138.

7. METALLIC CORROSION, PASSIVITY AND PRO-

tection (book), **U. R. Evans.** E. Arnold, London, England, 1937.

8. Cooling Mercury Arc Rectifiers, **H. O. Winograd.** *Allis-Chalmers Electrical Review*, Milwaukee, Wis., vol. 2, Sept. 1937, pp. 25–32.

9. The Electrochemistry of Corrosion, **R. H. Brown, R. B. Mears.** *Transactions*, Electrochemical Society, New York, N. Y., vol. 74, 1938, p. 495.

10. The Water-Cooled Steel-Tank Rectifier Corrosion Problem, **Emil J. Remscheid.** *AIEE Transactions* (*Electrical Engineering*), vol. 60, April 1941, pp. 173–78.

11. Protecting Cooling Systems Subject to Electrolysis, **Dr. E. W. Balis, Dr. H. A. Liebhafsky.** *General Electric Review*, Schenectady, N. Y., vol. 51, Sept. 1948, pp. 40–45.

12. Corrosion Control for Cooling Systems of Electric Apparatus, **E. J. Remscheid.** *General Electric Review*, Schenectady, N. Y., vol. 54, March 1951, pp. 36–40.

13. Controlling Factors in Galvanic Corrosion, **W. A. Wesley.** *Proceedings*, American Society for Testing Materials, Philadelphia, Pa., vol. 40, 1940, pp. 690–702.

14. Some Observations of the Potentials of Metals and Alloys in Sea Water, **F. L. LaQue, G. L Cox.** *Ibid.*, pp. 670–89.

15. The Practical Problems of Corrosion Vs. Corrosion and Protection at the Contacts of Dissimilar Metals, **U. R. Evans.** *Journal*, Society of Chemical Industry, London, England, vol. 47, 1928, p. 337.

16. Contact Corrosion of Pure Aluminum With Different Metals, Particularly Stainless Steel, **W. Geller.** *Korrosion und Metallschutz*, Berlin, Germany, vol. 15, Sept. 1939, p. 298.

17. Power Rectifiers, **E. J. Remscheid.** *General Electric Review*, Schenectady, N. Y., vol. 47, March 1944, pp. 25–28.

18. Electrode Behavior of Corroding Metals in Aqueous Solutions (book), **O. Gatty, E. C. Spooner.** Daredon Press, Oxford, England, 1938.

19. Dissimilar Metal Couples—Corrosion Resistance and Current Potential Measurements, *Industrial and Engineering Chemistry*, Easton, Pa., vol. 31, 1939, p. 608.

20. The Industrial Utility of Public Water Supplies in the United States. *Supply Paper No. 658*, United States Department of the Interior, Washington, D. C.

21. Performance of Pumped Ignitron Rectifiers, **C. C. Herskind, E. J. Remscheid.** *AIEE Transactions*, vol. 67, pt. I, 1948, pp. 215–18.

22. Operation and Maintenance of Mercury Arc Rectifiers, **R. O. Nixon.** *Allis-Chalmers Electrical Review*, Milwaukee, Wis., vol. 3, June 1938, pp. 8–13.

23. Preventing Rectifier Corrosion, **D. W. Borst.** *Coal Age*, New York, N. Y., vol. 52. Dec. 1947, pp. 89–93.

24. Corrosion Causes and Prevention (book), **F. N. Speller.** McGraw-Hill Book Company, Inc., New York, N. Y., 1935.

25. Corrosion Resistance of Metals and Alloys (book), **R. J. McKay, R. Worthington.** Reinhold Publishing Corporation, New York, N. Y., 1936.

26. Cathodic Protection (book). National Association of Corrosion Engineers, Houston, Tex., 1949.

27. Protective Films on Metals (book), **E. S. Hodges.** D. Van Nostrand Company, Inc., New York, N. Y., 1932.

28. The Fundamentals of Electrochemistry and Electrodeposition (book), **S. Glasstone.** American Electroplaters Society, Paterson, N. J., 1948.

29. Chemical Treatment to Mitigate Corrosion, **J. G. Jewell.** *Corrosion*, Houston, Tex., vol. 8, March 1952, pp. 100–08.

30. Basic Heat-Transfer Data in Electron Tube Operation, **Mrs. B. O. Buckland.** *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 1079–85.

31. Laminar-Flow Heat Transfer Coefficients for Ducts, **R. H. Norris, D. D. Streid.** *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 62, Aug. 1940, pp. 525–33.

No Discussion

Frequency Generating Equipment for Million Watt Navy Transmitter

D. G. ROBERTSON
MEMBER AIEE

Synopsis: Novel methods are used to obtain crystal stability and still permit continuous coverage of the frequency range. The type *o-146/FRT-3* Frequency Generator-Monitor described is part of the United States Navy's new model *AN/FRT-3* superpower very-low-frequency transmitter.

THE DESIGN OF frequency generating equipment for a 1000-kw Navy transmitter required the application of new circuits and combinations of circuits to meet the following general specification requirements:

1. The generation of a radio-frequency signal continuously variable from 14.5 to 35.0 kc, with a frequency stability of 0.001 per cent;
2. An on-off keying capability of speeds up to 75 cycles per second without the use of relays when polar or neutral keying signals are supplied;
3. A frequency-shift keying capability, either polar or neutral, with deviations up to plus and minus 25 cycles, at speeds up to 24 cycles per second.

The equipment design to meet these requirements is called the Frequency Generator-Monitor. All components necessary for the proper operation of the frequency generator-monitor are contained in 12 units mounted in the 3-bay rack-type enclosure shown in Fig. 1.

Basic equipment consists of two variable-frequency sources, a master-oscillator, and a monitor. Auxiliary equipment is also supplied to compare the two frequencies. This consists of a deviation indicator and an automatic frequency control. Elements mounted in the left bay and the upper unit of the center bay comprise the master oscillator. The second and third units in the center bay are the deviation indicator and automatic-frequency-control units respectively. The remaining bay constitutes the monitor. Fig. 2 is a block diagram showing the relationships of these four functions.

The master oscillator is comprised of five units:

1. The variable frequency oscillator;
2. The mixer amplifier and keyer;
3. Number 1 and number 2 crystal oscillators;
4. The master oscillator power supply;
5. The frequency shift meter.

Block diagram Fig. 3 shows the relationships of these units and the output and input information associated with each unit.

Radio-frequency voltage to drive the transmitter is generated by heterodyning the 214.5- to 235.0-kc output of the variable-frequency oscillator with the output of crystal oscillators no. 1 or no. 2, and using the difference frequency. The 200-kc output of oscillator no. 1 is used for on-off keying. On-off keying is accomplished without relays by using the keying signals as a direct control on the signal to the transmitter. The 200-kc output of oscillator no. 2, modified by the reactance tube up to plus or minus 25 cycles per second, is used for frequency-shift keying. Relays are used in frequency-shift keying. The keying signals operate the relays and the relay contacts apply control voltage to the reactance-tube grids. The amount of shift is indicated on the frequency-shift meter which compares the frequency of crystal oscillator no. 2 against that of crystal oscillator no. 1.

The monitor also consists of five units:

1. The fixed frequency divider;
2. The frequency generator;

Paper 53-198, recommended by the AIEE Radio Communication Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted December 26, 1951; made available for printing March 30, 1953.

D. G. Robertson is with the Radio Corporation of America, Camden, N. J.

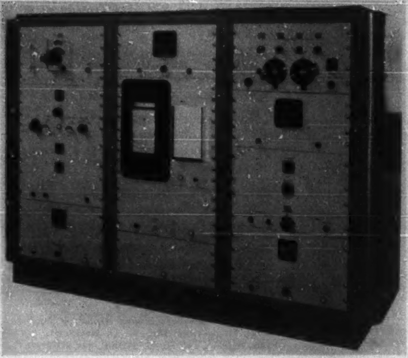


Fig. 1. Frequency generator-monitor

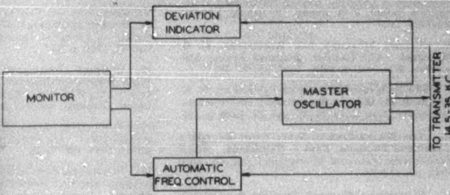

DEVIATION INDICATOR
MONITOR
MASTER OSCILLATOR
TO TRANSMITTER 14.5-35 KC.
AUTOMATIC FREQ CONTROL

Fig. 2. Block diagram, frequency generator-monitor

3. The frequency increment meter;
4. Crystal oscillators no. 3 and no. 4;
5. The monitor power supply.

Block diagram Fig. 4 illustrates the relationships of these units and pertinent output and input frequencies. The monitor generates its output frequency in an unusual manner, enabling it to maintain a stability within the specified tolerance of 0.001 per cent. It is accomplished through the use of a bridge-stabilized oscillator[1] in conjunction with a binary-type variable-frequency divider. The bridge-stabilized oscillator, crystal oscillator no. 3, provides stability, and the variable-frequency divider permits frequency variation. Fig. 5 illustrates this action. The 200-kc frequency generated by crystal oscillator no. 3 is passed through the fixed-frequency divider to obtain a reference frequency of 100 cycles. The frequency generator, by virtue of its variable-frequency divider, then acts to produce the monitor output signal having the stability of this 100 cycle signal.

The basic monitor frequency is generated by a simple Colpitts oscillator. This oscillator is continuously variable over the range of 645 to 850 kc. The output of this oscillator is fed to the variable-frequency divider, which can be adjusted to divide by any whole number from 6,450 to 8,500 inclusive. For example, if the variable oscillator frequency is 645,000 cycles per second, and the divider is set to divide by 6,450, the output frequency is 100 cycles per second. This signal and the 100 cycle signal derived from crystal oscillator no. 3 are fed into a phase detector. The output of the phase detector acts on a reactance tube coupled to the variable oscillator in such a manner as to compensate for any tendency to drift on the part of the variable oscillator. Thus, the variable oscillator is "locked-in" with crystal oscillator no. 3. By this method frequencies with crystal stability are generated at 100-cycle intervals throughout the range of 645 to 850 kc.

This voltage is fed to a fixed divider having a division ratio of ten. This gives an output of 64.5 to 85 kc, in steps of ten cycles. To derive the final monitor output frequency, this signal is heterodyned with a 50-kc signal derived from crystal oscillator no. 3. The difference frequency of 14.5 to 35 kc, in 10 cycle steps, is thus obtained. In order to obtain continuous-frequency coverage, crystal oscillator no. 4 is switched in place of crystal oscillator no. 3 in the final heterodyne circuit. Oscillator no. 4 is provided with a vernier adjustment to lower its frequency 40 cycles from 200 kc. This results in increments of up to ten cycles in the monitor output frequency. The frequency-increment meter is provided to measure the frequency difference between oscillators 3 and 4. It is calibrated 0 to 10 cycles to be in accord with the monitor output.

The automatic-frequency-control unit, Fig. 5, samples the output signals of both the master oscillator and the monitor. It compares the signals, and when any frequency difference exists, it produces 2-phase power. This power operates a synchronous motor to turn a variable capacitor in the master oscillator's variable-frequency oscillator. By this means, the master-oscillator frequency is continually corrected to give it the stability of the monitor.

Of particular interest is the variable-frequency divider. It consists of a chain of 16 driven flip-flop binary dividers. Each binary consists of two triodes connected to resemble a multivibrator, so that when one triode is conducting, the other is cut off. The constants are so chosen that the circuit is stable in either state of conduction, and the application of a triggering pulse will reverse the cir-

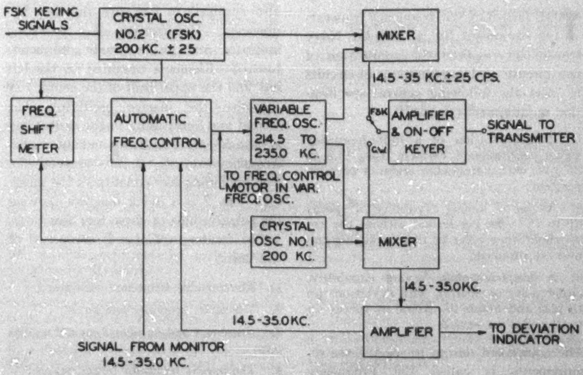

FSK KEYING SIGNALS
CRYSTAL OSC. NO.2 (FSK) 200 KC. ± 25
MIXER
14.5-35 KC.±25 CPS.
FREQ. SHIFT METER
AUTOMATIC FREQ. CONTROL
VARIABLE FREQ. OSC. 214.5 TO 235.0 KC.
FSK
CW
AMPLIFIER & ON-OFF KEYER
SIGNAL TO TRANSMITTER
TO FREQ. CONTROL MOTOR IN VAR. FREQ. OSC.
CRYSTAL OSC. NO. 1 200 KC.
MIXER
14.5-35.0KC.
14.5-35.0KC.
AMPLIFIER
TO DEVIATION INDICATOR
SIGNAL FROM MONITOR 14.5-35.0 KC.

Fig. 3. Block diagram, master-oscillator units

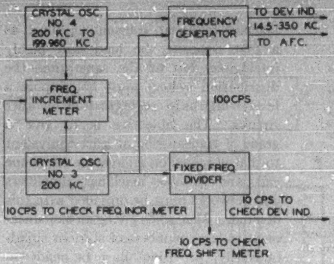

CRYSTAL OSC. NO. 4 200 KC. TO 199.960 KC.
FREQUENCY GENERATOR
TO DEV. IND.
14.5-350 KC.
TO A.F.C.
FREQ. INCREMENT METER
100 CPS
CRYSTAL OSC. NO. 3 200 KC.
FIXED FREQ. DIVIDER
10 CPS TO CHECK FREQ. INCR. METER
10 CPS TO CHECK DEV. IND.
10 CPS TO CHECK FREQ. SHIFT METER

Fig. 4. Block diagram, monitor units

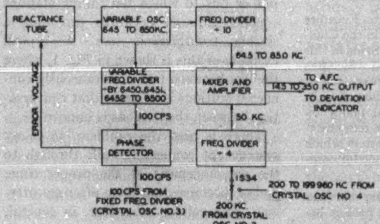

REACTANCE TUBE
VARIABLE OSC. 645 TO 850 KC.
FREQ. DIVIDER ÷ 10
64.5 TO 85.0 KC.
VARIABLE FREQ. DIVIDER BY 6450, 6451, 6452 TO 8500
MIXER AND AMPLIFIER
TO A.F.C.
14.5 TO 350 KC OUTPUT
TO DEVIATION INDICATOR
ERROR VOLTAGE
100 CPS
50 KC.
PHASE DETECTOR
FREQ. DIVIDER ÷ 4
100 CPS
100 CPS FROM FIXED FREQ. DIVIDER (CRYSTAL OSC. NO. 3)
1S34
200 TO 199.960 KC FROM CRYSTAL OSC NO 4
200 KC. FROM CRYSTAL OSC. NO. 3

Fig. 5. Block diagram, frequency-generator unit

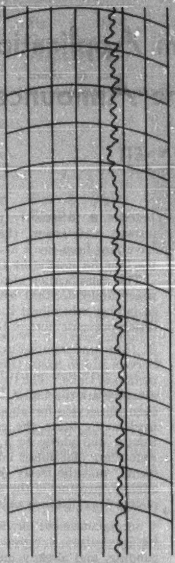


Fig. 6. Master-oscillator frequency stability, with automatic frequency control. Horizontal scale is 0.4 cycles per division. Vertical scale is 6.0 minutes per division

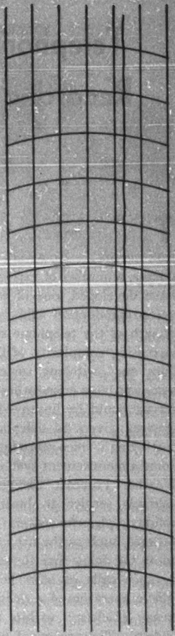


Fig. 7. Master-oscillator frequency stability without automatic frequency control. Horizontal scale is 5.0 cycles per division. Vertical scale is 6.0 minutes per division

cuit. The application of two pulses of the same polarity causes 1 cycle of operation, giving a divide-by-two circuit which is independent of frequency. Four of these binaries connected in series would ordinarily divide by 16, but when two feedback circuits are properly connected between stages a decade (divide-by-ten) circuit results. In this equipment the 16 binaries are arranged into three decades, followed by four binaries, resulting in a basic division of 10×10×10×16, or 16,000. It can be seen that even though each binary has but two conditions (one triode or the other conducting) there are 16,000 different combinations of these conditions, one for each of the 16,000 input pulses.

To make the chain into a variable divider, one of these conditions is taken as the zero-count point and whenever this particular condition exists a thyratron is fired. The output pulse of this thyratron is called the reset pulse, and is fed through four selector switches to one triode grid or the other of every binary in the dividing chain. If for instance, the selector switches are set to obtain a division of 8,500 the reset pulse is applied to the proper binary grids so that the dividing chain is immediately placed in the same condition as though it had received 7,500 input pulses. From this condition it takes only 8,500 input pulses to bring the dividing chain back to the zero condition. If the reset pulse is taken as the output, there will be one output pulse for every 8,500 input pulses. Under these conditions, an input of 850 kc to the divider will result in an output of 100 cycles. The proper setting of the selector switches permits the choice of any division ratio from 6,450 to 8,500 inclusive.

Fig. 6 shows the frequency-stability characteristics of the master oscillator when operating with automatic-frequency control. It is a representative 1-hour sample of the output frequency as recorded during a 6-hour heat run. Since the horizontal scale is 0.4 cycle per division, and the run was taken at a frequency of 22 kc, the total variation shown is 0.00032 per cent. The specifications require 0.001 percent. Fig. 7 is a sample of a similar frequency versus time run, but with the automatic frequency control feature disabled. In this case the horizontal scale is 5.0 cycles per division instead of 0.4 cycles. The total variation shown is 0.5 cycles, or 0.0023 percent of the 22-kc output frequency. It should be pointed out that the latter is ordinarily considered to be a high degree of frequency stability for transmitters operating in the very-low-frequency band. The master-oscillator output which is used to drive the transmitter is obtained by heterodyning a one-tube variable-frequency oscillator against a one-tube crystal oscillator. To generate the monitor frequency requires the use of some 75 tubes and their associated circuits. Thus in the event the monitor fails, transmitter operation can still be maintained with a useful degree of frequency stability. Achievement of this practical design goal satisfies the important requirement of dependability, which is an essential military characteristic.

Reference

1. The Bridge-Stabilized Oscillator, L. A. Meacham. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 26, Oct. 1938, p. 1278

◆

No Discussion

Telephone-System Applications of Recorded Machine Announcements

W. BENNETT
MEMBER AIEE

THE BUSINESS of the Bell System is primarily to provide communication services. However, as in any industry, a number of very useful side lines have developed, one of which is in the voice-reproduction field. With the vast growth of the telephone system and the increasing complexity of its networks of lines and switching systems, numerous instances have arisen wherein the general service could be improved or additional services given to telephone users by employing voice-reproduction devices. Some announcement systems, such as the Weather Forecast bureau, provide invaluable service to business men and holiday pleasure-seekers alike. Other systems, such as the use of a machine to quote the delay time for setting up long-distance calls in busy holiday seasons, relieve operators of a tiring task and increase efficiency within the telephone system. The process of providing these added services has, in turn, led to extensive development effort in the recording field by the Bell Telephone Laboratories and an appreciable investment has been made in announcement devices and the associated control equipment.

The applications of voice-recording equipment may be divided into the following four general groups, covered in greater detail below:

1. Announcements Made Direct to and Providing a Distinct Service to Subscribers: The two principal services in this group are Weather Forecast and "Time-of-Day."

2. Announcements Used to Assist Subscribers in Connection With Telephone Service: In this particular group, the principal field of application is in connection with certain Intercept announcements. It is inevitable that some telephone calls will reach vacant or disconnected terminals and it is undesirable, from the standpoint of both the calling subscriber and the telephone company, to permit these calls to be unanswered, since this would waste the time of the subscriber and tie up switching equipment. Therefore it has generally been the practice, until recently, to have an operator answer, or intercept, these calls. Facilities have now been made available to employ a machine for answering these calls.

This classification also includes the emergency announcement system which is used to inform subscribers if some unusual condition prevents normal telephone service.

3. Announcements Employed to Expedite Service and to Assist Operators in the Completion of Calls: The two main fields in this category are the Call Announcer, which has been in service for about 23 years and is used to facilitate the completion of calls from a dial phone to a nondial phone by converting dial pulses into voice announcements; and the Delay Quote system which advises operators of the delay time for completing long distance calls.

4. Announcement or Recording Services for Specialized Purposes: This field is expanding, and already includes price-quotation systems, ticket-reservation systems, various announcement systems, and business aids.

Types of Announcement Equipment

There are three types of voice-reproduction equipment in common use at the present time:

1. The optical method, utilizing a photographic-film record, and having an associated scanning system with photoelectric cells.

2. The magnetic method, using a magnetic pattern set up in magnetic media such as wires, tapes, drums, etc.

3. The mechanical method, in which the recording is engraved or embossed on a recording material.

Careful studies have been made before selecting announcement devices for each application, the optical and magnetic methods having been given the most extensive use in telephone-announcement work.

General Features

Each new announcement system has its individual features and has introduced its own individual problems but, in addition, there are several features common to all.

The transmission considerations on announcement systems are, of course, highly important. Because the speech on announcement systems is one way and the calling subscriber cannot ask the machine to repeat, transmission must, as a first requisite, be loud and highly intelligible, but at the same time the distribution system must not impose a level so high as to create crosstalk on other circuits. The objective is to give a grade of transmission at least equivalent to that obtained on an average telephone call, and suitable machines, amplifiers and networks must be provided with this in mind. Also, when giving an announcement to large numbers of stations simultaneously, provision must be made to prevent conversation between the connected stations. This is accomplished by the use of a low common-impedance output or bus to which the calling stations connect through networks providing a sizable transmission loss. A typical example of this is shown in Fig. 1, where the loss between two simultaneously connected stations is so great that conversation between the stations is impossible.

There is also the problem, in many systems, of connecting calls through to the announcement at the proper time, and of disconnecting them after an interval. The calling subscriber, in general, is allowed to hear the announcement for a long enough time to permit him to absorb the message but, on the other hand, a cutoff interval must be established to avoid prolonged listening which monopolizes valuable trunk and switching facilities. This problem is solved by a suitable switching circuit usually controlled by signals from the announcement machine.

Due to the importance of machine announcements, complete reliability of service is a primary consideration and therefore extensive alarm equipment is provided. Since the announcement machine is the heart of the system, it is the practice to furnish a standby machine with automatic transfer arrangements in the event of failure of the machine serving the load. Arrangements are also generally provided to permit an operator to make announcements if all machines fail.

Description of Systems

WEATHER FORECAST SERVICE

The weather announcement system has proved to be a very popular service since the first installation in 1939. In New York City, for example, the service averaged 71,000 calls per day in 1951. This rate varies extensively and has reached a daily high of about 270,000 calls under severe weather conditions.

The forecast information for this announcement system is received through the co-operation of the Weather Bureau. The information is transmitted over a

---

Paper 53-219, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted July 7, 1952; made available for printing April 23, 1953.

W. Bennett is with the Bell Telephone Laboratories, Inc., New York, N. Y.

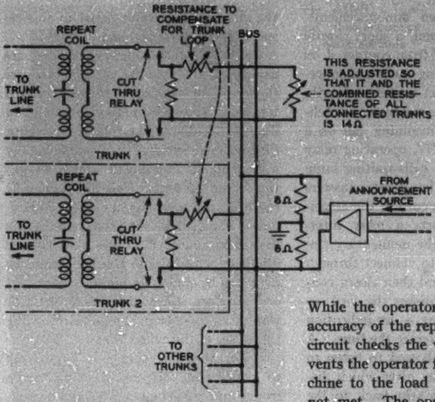

RESISTANCE TO COMPENSATE FOR TRUNK LOOP
REPEAT COIL
BUS
TO TRUNK LINE
CUT THRU RELAY
THIS RESISTANCE IS ADJUSTED SO THAT IT AND THE COMBINED RESISTANCE OF ALL CONNECTED TRUNKS IS 14Ω
TRUNK 1
REPEAT COIL
FROM ANNOUNCEMENT SOURCE
8Ω
8Ω
TO TRUNK LINE
CUT THRU RELAY
TRUNK 2
TO OTHER TRUNKS

Fig. 1 (left). Connection of announcement trunks to common bus

Fig. 2 (right). Weather - forecast-service recording room

direct teletype line to the recording room in a telephone central office. A typical recording room is shown in Fig. 2 with the teletype machine and recording turret. In addition to providing a convenient service to telephone subscribers, this system relieves the Weather Bureau personnel of considerable telephone-answering effort. It is customary to change the announcement at hourly intervals, or more frequently, if necessary, to take care of sudden weather changes.

This service presently uses magnetic-type announcing machines developed for the purpose.[1] The machines and amplifiers function on 110 volts a-c and, therefore, an inverter is provided for use if a power failure occurs. The inverter operates from a 48-volt battery in the telephone central office and produces 110 volts a-c.

The general arrangement of the announcement system is shown in Fig. 3. The trunks from various central offices are ordinarily terminated on the output bus at the main bureau but, if the area served is sufficiently large, subcenters are provided with two cable pairs between the main bureau and the subcenter.

As shown in Fig. 3, the output of the announcement machines is connected to the bus through pattern relays, which are designed to operate in a sequence giving each machine an equal amount of use in the load and standby positions. This pattern is automatically controlled and the proper machines are selected when an erase key in the turret is operated, preparatory to a new recording. While one machine is kept in service, the remaining two available machines are recorded simultaneously, but cannot be put in service until the operator has monitored both. While the operator monitors the verbal accuracy of the report, a speech monitor circuit checks the volume level and prevents the operator from connecting a machine to the load if specified limits are not met. The operator places the two newly recorded machines in their proper places as the load and standby machines by the operation of a single key. A speech-monitor circuit is connected across the bus as a check of the presence of speech and its volume level. If the volume is too low or if there is no speech, a transfer to the standby machine occurs. As an additional safeguard, such as the remote eventuality of all machines being out of service, the operator will give the announcement herself by plugging her microphone into a manual-announcement jack causing the operation of a pair of manual relays and the manual-service relay. A machine recording may be made during manual announcement by way of contacts *A* shown in Fig. 3.

As stated above, subcenters may be provided, with two cable pairs to the main bureau. Each pair has an amplifier at the subcenter, and the outputs of the two amplifiers are tied together through a bridge circuit, one arm of which is the subcenter bus. If either channel from the main bureau fails, the other can carry the load at the subcenter with a very slight drop in volume. Normally-operated relays are connected across the cable pairs, which will release and cause an immediate alarm if an open line should occur. A relay arranged to operate on speech is connected across the output of each amplifier under control of interrupters and, if the relays fail to operate on each interrupter cycle, an alarm is sounded. The bridge connection between the two channels prevents the output of one working channel from operating the speech-alarm relay of the other channel, if it is out of order.

The announcement given for weather is approximately 25 seconds in length and, therefore, incoming calls are connected through immediately, at any part of the announcement. However, a distributing and timing circuit is provided to control supervision signals in the incoming trunk circuits. After several announcements, the incoming trunk is arranged to send a disconnect signal toward the originating central-office equipment.

TIME ANNOUNCEMENT

Correct time has become a very important item to most of us. Early in telephone history, subscribers acquired the habit of calling the operator for the correct time. This became burdensome to the operator, particularly in the larger areas, and for this reason a number of these areas were arranged with a centralized time-of-day bureau that could be reached by dialing a listed number. In this system, one operator at the central point announced the time at 15-second intervals. This repeated announcement was tedious and operators tired after a short time. Due to the repetitive nature of time announcements, it appeared that the use of a machine announcement system would materially contribute to the convenient and efficiency of the operating personnel and accordingly, it was made available. The service is extensively used. In New York City, the average daily calling rate for 1951 was 68,000. During the annual baseball World Series, the time bureau also carries the inning-by-inning score and in 1951, 1,456,000 calls reached the bureau during the six days of the series.

The machine pictured in Fig. 4 is built by the Audichron Company, Atlanta, Ga., and is presently furnished for time service. The unit shown is equipped with two machines arranged for automatic transfer in the event of failure. However, a single machine has also been made available for use in towns with a small amount of time-of-day traffic.

The machine, which uses the optical method of voice reproduction, is arranged to give a statement such as: "At the tone, the time will be one twenty-two and ten seconds," followed by a short spurt of tone. The announcements are given at 10-second intervals. The machines are driven by 60-cycle commercial power, when it is considered to be regulated closely enough. In other installations, a local power source with closely controlled frequency can be provided. As in the weather service, inverters operating from the 48-volt battery supply 110-volts a-c if a power failure occurs.

The distributing network and subcenter arrangement is similar to that shown in Fig. 3 for that of the Weather Bureau. The general arrangement of a 2-channel time-bureau installation is schematically illustrated in Fig. 5. Here, one announcing machine serves the load, with the other acting as a standby.

If both announcement devices should fail, a visual and audible alarm is given, and a key is operated to transfer the load to a manual-announcement source which employs a third amplifier, not associated with the machines. These announcements may be made from either a microphone and emergency key and lamp box, or from a regular switchboard position, as indicated in Fig. 5. The operator, of course, obtains the time from a clock and the tone spurt is not given.

Inasmuch as a time announcement is short, incoming calls are connected through only between announcements. With machine operation, a short closure of ground on lead *RV* from contact *A* on the machine, preceding each announcement, serves to connect waiting subscribers to the bus. During emergency switchboard operation, an incoming call on a trunk grounds lead *CW*, operating relay *CW* and lighting the calls-waiting lamp in the switchboard. The operator answers the call by plugging a cord in the associated jack, which starts a relay interrupter. This causes the momentary closure of contacts *RV* to connect through the waiting trunks, and then closes contact *AN* which extinguishes the cord supervisory lamp, serving as an indication to the operator to announce the time. A calling subscriber is allowed to listen to several announcements, after which disconnect supervision is sent back to the originating equipment.

Various alarms are provided at the main bureau. These include voice alarms, fuse alarms, faulty distributing-relay operation, and a current-interruption alarm which operates if there is a momentary failure of the 110-volt a-c power, in order to indicate that the time may be slightly off.

Recorded Intercept Announcements

This service is one of the newer users of recorded announcements and will reduce the amount of operator effort in this respect by about 50 per cent.

Calls in the telephone system which are usually intercepted by an operator are those that reach vacant or disconnected terminals, changed number terminals, or lines, in trouble. The calls to vacant or disconnected terminals can be handled by a machine announcement, which advises the calling subscriber that the number he reached has been disconnected and requests him to make sure that he is calling the right number. Calls to changed numbers or to lines in trouble cannot be answered by such a simple statement, and are therefore still routed to an operator who gives definite information regarding the status of the called line.

It has been the practice in large areas to concentrate the intercepting traffic from a number of offices to a single switchboard or desk. Fig. 6 is a functional schematic of the arrangements usually furnished at a central intercepting point, and the trunk equipment at both ends of one trunk from an outlying office to the central point. At the outlying office, the calls to be intercepted are directed to the intercepting trunk by the switching equipment, which also sends signals ahead to indicate the type of answer to be given at the central point. Calls to vacant or disconnected terminals will reach the announcement machine, calls

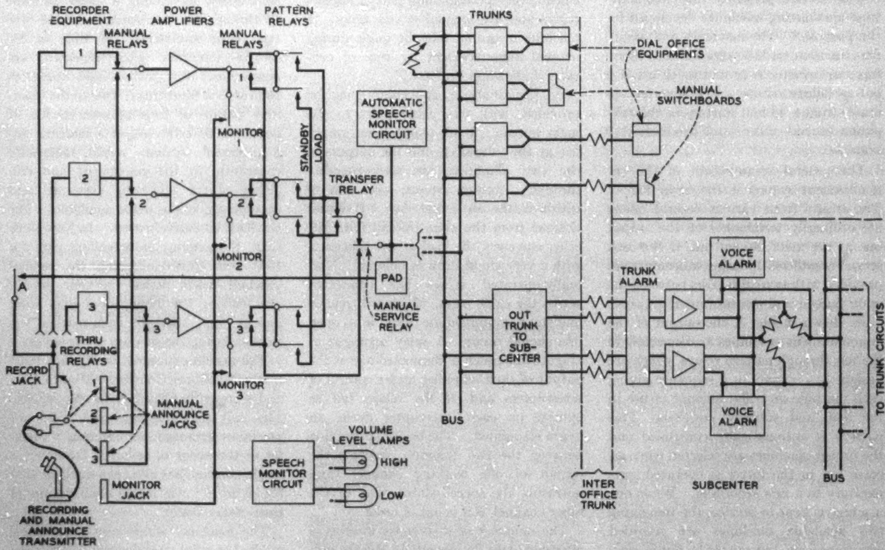

RECORDER EQUIPMENT
POWER AMPLIFIERS
PATTERN RELAYS
MANUAL RELAYS
MANUAL RELAYS
1
2
3
MONITOR 1
MONITOR 2
MONITOR 3
STANDBY
LOAD
TRANSFER RELAY
AUTOMATIC SPEECH MONITOR CIRCUIT
PAD
MANUAL SERVICE RELAY
A
THRU RECORDING RELAYS
RECORD JACK
MANUAL ANNOUNCE JACKS
MONITOR JACK
RECORDING AND MANUAL ANNOUNCE TRANSMITTER
SPEECH MONITOR CIRCUIT
VOLUME LEVEL LAMPS
HIGH
LOW
TRUNKS
DIAL OFFICE EQUIPMENTS
MANUAL SWITCHBOARDS
BUS
OUT TRUNK TO SUB-CENTER
INTER OFFICE TRUNK
TRUNK ALARM
VOICE ALARM
VOICE ALARM
SUBCENTER
BUS
TO TRUNK CIRCUITS

Fig. 3. Functional schematic of weather-bureau network

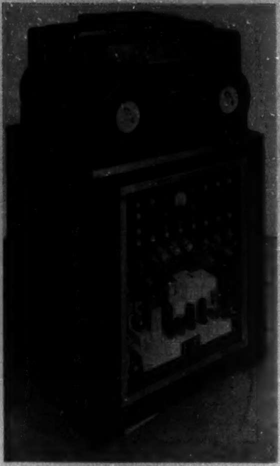


Fig. 4 (left). Typical time-announcement machine

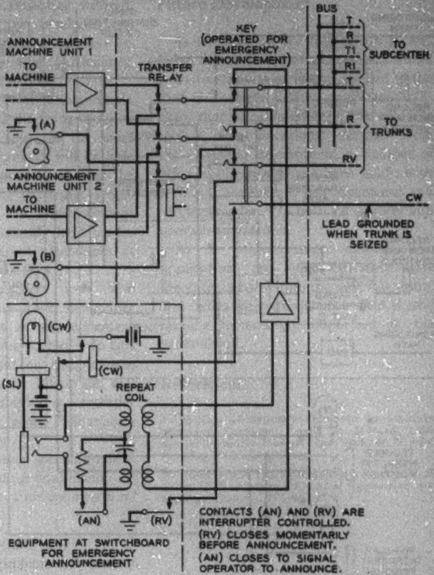

BUS
KEY (OPERATED FOR EMERGENCY ANNOUNCEMENT)
T
R
T1
R1
TO SUBCENTER
ANNOUNCEMENT MACHINE UNIT 1
TO MACHINE
TRANSFER RELAY
T
R
TO TRUNKS
RV
(A)
ANNOUNCEMENT MACHINE UNIT 2
TO MACHINE
CW
LEAD GROUNDED WHEN TRUNK IS SEIZED
(B)
(CW)
(CW)
(SL)
REPEAT COIL
(AN)
(RV)
EQUIPMENT AT SWITCHBOARD FOR EMERGENCY ANNOUNCEMENT
CONTACTS (AN) AND (RV) ARE INTERRUPTER CONTROLLED.
(RV) CLOSES MOMENTARILY BEFORE ANNOUNCEMENT.
(AN) CLOSES TO SIGNAL OPERATOR TO ANNOUNCE.

Fig. 5 (right). Functional schematic of time of day bureau

to lines in trouble will reach an operator's jack and lamp for this service, while calls to changed number terminals will reach another jack and lamp. Referring to Fig. 6, a call to a changed number will cause relay *TRO* of the outgoing trunk to operate and connect battery and ground toward the incoming trunk, but the polarity is such that relay *P* will not operate, and the call is routed to the operator's Regular answering jack and lamp. A call to a line in trouble will operate both relays *TRO* and *RV*, reversing the polarity and causing polarized-relay *P* to operate and direct the call to the operator's Trouble lamp and jack. Calls to either vacant or disconnected terminals cause the momentary operation of relay *M* of the outgoing trunk, applying a 115-volt battery to both sides of the line, which operates the simplex-connected relay *SX* at the incoming end. The operation of relay *SX* sends the call to the announcement machine.

The announcement machines are so constructed that contact *CT* closes during the period between announcements to operate the *CT* relays of the trunks. The pulses from contact *CT* also cause the operation of a counting circuit and, after a predetermined number of announcements, the call is transferred to the operator's Regular jack and lamp, on the assumption that the calling party requires some information in addition to the machine announcement.

The machines used for intercepting service are the magnetic type and are furnished in duplicate, with automatic transfer. In addition to advising the calling subscriber that he reached a disconnected number, he is told that he is listening to a recorded announcement. This latter statement is necessary, since observations have shown that without it, about 40 per cent of the calling subscribers attempt to converse with the machine, believing it to be an operator.

EMERGENCY ANNOUNCEMENT SYSTEM

On rare occasions, magnetic-type announcement machines have been used to advise telephone subscribers if an emergency, such as an air-raid alert, causes the temporary discontinuance of operator service. The method of applying this announcement is interesting due to its simplicity.

In a manual (nondial) central office, the announcement is applied by superimposing it on the battery feed to the line relays.

In any dial central office, some operator services are required, and therefore trunk circuits are furnished from the dial equipment to these operators. Since subscribers are accustomed to receiving audible ringing signals between the end of dialing and the answer of the called party, the trunk circuits have provision for applying audible-ringing signals toward the calling party, through a low capacity condenser (usually 0.02 to 0.04 microfarad), until such time as the call is answered. Under an emergency condition, the recorded announcement is applied by simply disconnecting the ringing from a group of trunks, and transferring the common leads to the announcement bus.

CALL ANNOUNCER

The introduction of dial-telephone systems could not be accomplished in all central offices at the same time. This situation called for some arrangement for completing calls from dial to manual telephones and, accordingly, an arrangement known as the call indicator was originally provided at the manual office. This system was such that the dialed number would appear before the manual operator as lighted numerals. With further expansion of the telephone system and the introduction of large tandem centers using common control equipment, the call announcer was developed for use at these tandem centers. This machinery brought about an appreciable reduction of equipment and modification needed at the manual office.

If a call is made from a dial area to a line in a manual office, the number dialed is recorded by relay operations. When the operator in the distant manual office is ready to serve the call, the operated relays cause the machine to announce the

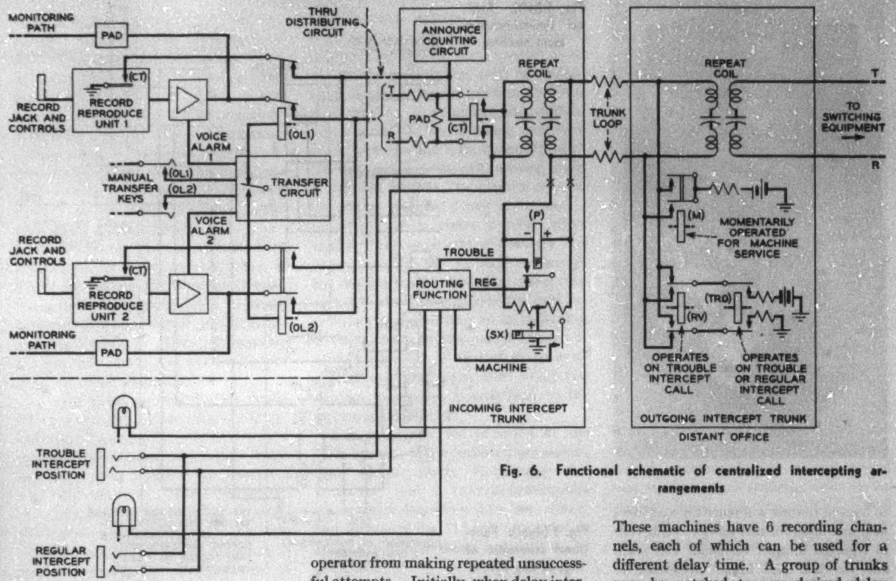

MONITORING PATH
PAD
RECORD JACK AND CONTROLS
(CT)
RECORD REPRODUCE UNIT 1
VOICE ALARM 1
(OL1)
MANUAL TRANSFER KEYS
(OL1)
(OL2)
TRANSFER CIRCUIT
VOICE ALARM 2
RECORD JACK AND CONTROLS
(CT)
RECORD REPRODUCE UNIT 2
(OL2)
MONITORING PATH
PAD
THRU DISTRIBUTING CIRCUIT
T
R
PAD
ANNOUNCE COUNTING CIRCUIT
REPEAT COIL
(CT)
TRUNK LOOP
(P)
TROUBLE
ROUTING FUNCTION
REG
(SX)
MACHINE
INCOMING INTERCEPT TRUNK
REPEAT COIL
T
TO SWITCHING EQUIPMENT
R
(M)
MOMENTARILY OPERATED FOR MACHINE SERVICE
(TRO)
(RV)
OPERATES ON TROUBLE INTERCEPT CALL
OPERATES ON TROUBLE OR REGULAR INTERCEPT CALL
OUTGOING INTERCEPT TRUNK
DISTANT OFFICE
TROUBLE INTERCEPT POSITION
REGULAR INTERCEPT POSITION

Fig. 6. Functional schematic of centralized intercepting arrangements

digits of the number, in proper sequence, to the distant operator.

The call announcer machine itself is of the optical type, and can announce the digits 0 to 9 and party line letters *W*, *R*, *J*, and *M*. Two machines are ordinarily in use, each carrying half the load, but in the event of a speech failure on one of the machines, an automatic transfer connects all the load to one machine.

Delay Quotations for Toll Calls

As mentioned in the introduction, another newcomer to the machine announcement field is the Delay-Quotation system. This was developed to relieve operators of the repetitious chore of quoting toll-line delay times during periods of heavy traffic such as Mother's Day, Christmas, etc.

The general arrangement for handling these quotations on delayed toll calls is shown in Fig. 7. Referring to this figure, assume that an operator at point *A* dials a call for point *C* via a toll crossbar office at point *B*, and that the switching circuits, testing, find that all trunks from *B* to *C* are busy. If the delay is to be of such length that the usual supervisory signals will not adequately inform the originating operator of the availability of trunks to point *C*, it becomes desirable to quote a period of time before the call can be completed, to prevent the originating operator from making repeated unsuccessful attempts. Initially, when delay intervals became 1/2-hour or more, the delay quotation was handled by patching (at the supervisory panel) the overflow control circuit to a jack appearing before an operator who gave the quotation, but since the development of the machine system, the operators are no longer required to do this. After the patch to the delay-quote jack has been set up, the calls will be directed to overflow trunks, the originating operator hears the quotation, and turns the call over to her own delayed-call operator for completion.

The jacks at the supervisory panel that carry the delay quotations are connected to magnetic-type recorders. These machines have 6 recording channels, each of which can be used for a different delay time. A group of trunks may be patched to any desired delay announcement. Contacts on the machine are provided to give supervisory flashing signals toward an outward operator, before the announcement occurs, to attract her attention in case she has left the connection.

No standby machines are furnished on these installations but, if a voice failure should occur, an alarm is given and the operator can quote the delays.

Specialized Applications

Voice-recording and announcement machines have been put in service for a number of uses as a convenience or important business aid in furnishing particular serv-

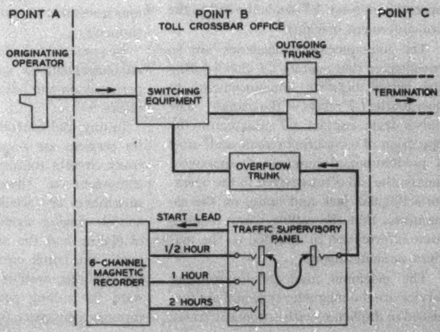

POINT A
POINT B
TOLL CROSSBAR OFFICE
POINT C
ORIGINATING OPERATOR
OUTGOING TRUNKS
SWITCHING EQUIPMENT
TERMINATION
OVERFLOW TRUNK
START LEAD
TRAFFIC SUPERVISORY PANEL
6-CHANNEL MAGNETIC RECORDER
1/2 HOUR
1 HOUR
2 HOURS

Fig. 7. Diagram of delay-quote system

ices. Several typical applications are described in the following paragraphs.

One of the earliest applications of magnetic type recorders for an announcement system was made in New Jersey in 1937 to give potato growers up-to-the-minute information regarding prices, car movements, loadings, and holdings. To receive this information it was necessary to call a listed number and, since the installation was in manual central-office area, the operator completed the call via a jack to the announcement bus. The system employed two machines, one in service, and the other being held as a spare or for new recordings.

A management announcement system, used on a few trial installations, permits the management of large corporations to disseminate important information to certain groups of employees without these employees leaving their regular work locations. The information is recorded on a magnetic-type machine and interested employees call a number in the company private-branch exchange to hear this announcement. Since it is often necessary to make a new recording without interrupting service, two machines are provided. These machines can also be operated sequentially if a particularly long announcement is needed.

It is planned to provide arrangements at the New York Stock Exchange to give price quotations on some of the most active stocks. The brokers will be able to obtain the latest price by dialing the number listed for the particular stock. The quotations on the announcement machines will be constantly kept up-to-date by recordings from attendants at the trading posts on the floor.

Magnetic-type recorders were also used in a large New York newspaper office in 1937, not as an announcement system, but as recording devices. In this instance, reporters out on assignments, could go to the nearest telephone, call their office and be connected to a recorder on which they would record their reports for later transcription.

There have been other special applications of voice recording equipment, including railroad-reservation information.[2]

Future Development

As far back as the late 1920's it was found that recorded announcements would serve to expedite service as applied to the Call Announcer. This experience indicated that voice recording was a valuable tool in communications. As seen from the foregoing descriptions, a number of other applications have been made with resulting advantages in convenience and services rendered to both subscribers and operating personnel.

At the present stage, various other applications are being explored, more are anticipated within the next few years, and indications are that this field should become increasingly important in the years ahead.

References

1. A NEW RECORDING MEDIUM FOR TRANSCRIBED MESSAGE SERVICES, James Z. Menard. *Bell System Technical Journal*, New York, N. Y., vol. 31, May 1952, pp. 530–40.

2. ELECTRONIC TICKET-RESERVING SYSTEM, W. R. Triem, J. S. Jammer. *Electrical Engineering*, vol. 69, Sept. 1950, pp. 775–80.

No Discussion

Heating of Fat-Muscle Layers by Electromagnetic and Ultrasonic Diathermy

HERMAN P. SCHWAN
ASSOCIATE MEMBER AIEE

EDWIN L. CARSTENSEN
NONMEMBER AIEE

KAM LI
NONMEMBER AIEE

IN CLINICAL practice, it is frequently desirable to produce heat in the muscular tissue which lies below subcutaneous layers of fat. Until recently short-wave diathermy provided the most satisfactory answer to the general problem of deep heating. However, biophysical research[1] has shown that, when electric currents pass through alternate layers of fat and muscle, the fatty tissue is selectively heated. Although the ratio of heat developed in fat to that in muscle becomes smaller as the frequency increases, even at the highest practical frequencies the ratio is still greater than unity. It was in part to overcome this difficulty that electromagnetic and ultrasonic radiation were introduced as forms of diathermy following the second World War. The heating of homogeneous tissues by radiant energy has previously been discussed.[2] However, the important fat-muscle problem as it applies to the new forms of radiation diathermy has not, up to now, been given quantitative consideration.

Radiation Diathermy

The discussions for heat development in the case of electromagnetic and acoustic radiation are completely analogous. In both cases, radiant energy flows through the tissue and is, in part, absorbed and converted into heat. At the interfaces separating two different tissues, reflections occur giving rise to standing waves which may materially influence the pattern of heat development. In order to obtain a complete analytical description of this problem it will only be necessary to assume knowledge of the complex propagation constants γ which characterize the various media. These propagation constants in turn, are either directly measurable quantities or can be simply related to measurable quantities. The quantitative conclusions of this analysis depend, therefore, upon measurements of the electromagnetic and acoustic properties of the tissues involved. Sources of these data are discussed in subsequent sections.

For the present discussion the conditions found in the body will be idealized, as shown in Fig. 1. Plane waves are assumed incident normally on plane and parallel interfaces, which separate the

---

Paper 53-206, recommended by the AIEE Electrical Techniques in Medicine and Biology Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 12, 1953; made available for printing April 2, 1953.

HERMAN P. SCHWAN, EDWIN L. CARSTENSEN, and KAM LI are with the Moore School of Electrical Engineering, and Department of Physical Medicine, Graduate School of Medicine, University of Pennsylvania, Philadelphia, Pa.

Aided in part by a grant from the National Foundation for Infantile Paralysis, Inc., and in part by the Aeromedical Equipment Laboratory, Philadelphia, Pa.

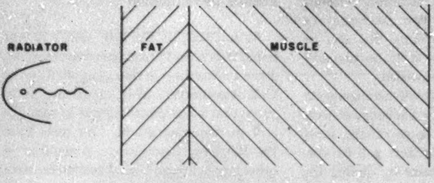


Fig. 1. One-dimensional representation of the fat-muscle configuration as found in the body

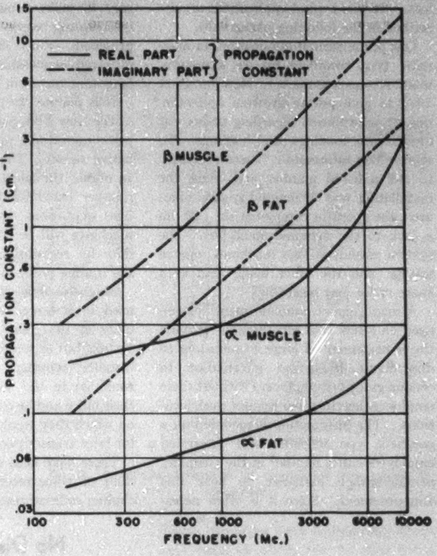


Fig. 2 (right). Electromagnetic propagation constant $\gamma = \alpha + j\beta$ for fat and muscle

layer of subcutaneous fat from the coupling medium (for example, water or air) on one side, and from the semi-infinite region of muscular tissue on the other side. The fat layer is taken to have a thickness *d;* the layer of muscle is assumed so thick that the waves transmitted into it are completely absorbed. This assumption is justified for most practical cases in view of the high absorption coefficient of muscle tissue, as shown in Figs. 2 and 4. The origin of the space co-ordinates is taken at the fat-muscle boundary and reference fields are defined at this point. For simplicity, the reference field E_0 will be taken as the effective incident wave at the fat-muscle boundary $x=0$. Of course, the effective incident wave is the sum of the initial wave transmitted directly from the generator and all positively directed reflection arising within the fatty layer. However, it can be shown[3] that the reflection factor for this sum is the same as for a single wave. Hence the complete picture can be given in terms of the effective incident wave.

In the following, the symbol E will be used for the electric field in the electromagnetic case, or for excess pressure in the acoustic case, while H will be used interchangeably for magnetic field or particle velocity. The field in fat then consists of incident and reflected waves E_1 and E_2.[4,5]

Table I. Dielectric Constant ϵ of Various Tissues at Different Frequencies

Tissue	100	300	1,000	3,000	10,000
	Frequency, Megacycles				
Muscle	74.0	60.0	54.0	54.0	45.0
Horse fat	8.0	7.1	5.6	4.4	3.3
Pork fat	5.2	4.6	3.4		

Table II. Resistivity ρ of Various Tissues at Different Frequencies

Tissue	100	300	1,000	3,000	10,000
	Frequency, Megacycles				
Muscle	100	100	86	45	8
Horse fat	1,550	1,350	1,100	850	380
Pork fat	3,800	3,150	2,450		

$$E_f = E_1 + E_2 = E_0(e^{-\gamma_f x} + p e^{\gamma_f x})$$

$$H_f = H_1 + H_2 = \frac{\gamma_f}{j\mu_f\omega}(e^{-\gamma_f x} - p e^{\gamma_f x})$$

where

μ = permeability in electromagnetic case and density in the acoustic case
p = the complex reflection factor of the fat-muscle boundary
γ = the propagation constant

The time factor $e^{j\omega t}$ is implied throughout. By applying the boundary conditions[4,5] it is seen that

$$p = \rho e^{j\varphi} = \frac{\dfrac{\gamma_f}{\mu_f} - \dfrac{\gamma_m}{\mu_m}}{\dfrac{\gamma_f}{\mu_f} + \dfrac{\gamma_m}{\mu_m}} \tag{1}$$

The subscripts f and m refer to fat and muscle, respectively. In the electromagnetic case, where it is reasonable to assume $\mu_f = \mu_m = 1$ this reduces to the following

$$p = \frac{\gamma_f - \gamma_m}{\gamma_f + \gamma_m}$$

The wave transmitted into the muscle is characterized by

$$E_m = (1+p)E_0 e^{-\gamma_m x}$$

$$H_m = (1-p)\,\frac{\gamma_m}{j\mu_m\omega}\,E_0 e^{-\gamma_m x}$$

The intensity I, or energy flow per unit area of the beam, is given by the complex product of electric and magnetic fields, complex Poynting's vector, in the electromagnetic case and by the complex product of pressure and particle velocity in the acoustic case.

Writing $\gamma = \alpha + j\beta$ and using $p = \rho e^{j\varphi}$, the intensity in fat is

$$I_f = \frac{1}{2}EH^* = \frac{1}{2}\,\frac{\beta_f + j\alpha_f}{\mu_f\omega}\,E_0E_0^* \times$$
$$[e^{-2\alpha_f x} - \rho^2 e^{2\alpha_f x} + j2\rho \sin(2\beta_f x + \varphi)]$$

The asterisk is used to signify the complex conjugate of the number.

The rate of development of heat per unit volume is given by the negative space rate of change of the real part of this intensity,[4] that is,

$$-\frac{\partial}{\partial x}\mathrm{Re}\, I_f = \frac{1}{2}\,\frac{2\alpha_f\beta_f}{\mu_f\omega}\,E_0^2 \times$$
$$[e^{-2\alpha_f x} + \rho^2 e^{2\alpha_f x} + 2\rho\cos(2\beta_f x + \varphi)] \tag{2}$$

In the electromagnetic case, this reduces to

$$\frac{1}{2}\kappa_f E_0^2[e^{-2\alpha_f x} + \rho^2 e^{2\alpha_f x} + 2\rho\cos(2\beta_f x + \varphi)]$$

where κ_f is the electrical conductivity of fat, since the product $\alpha_f\beta_f = (\mu_f\kappa_f\omega)/2$.

From the continuity of E across the

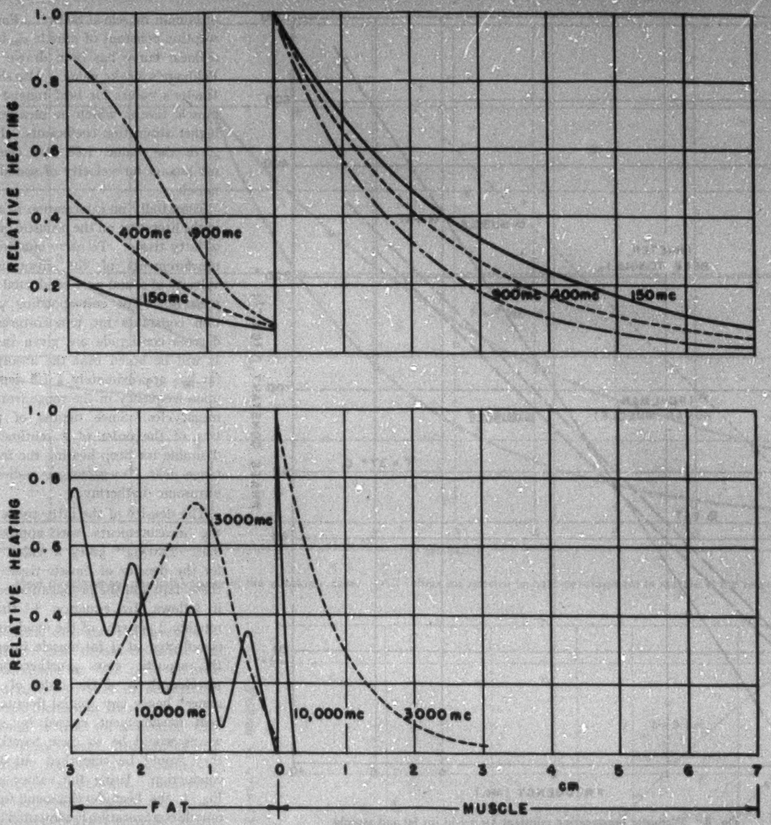


Fig. 3. Heat development per unit volume in electromagnetic case. All values are relative to heat development in muscle at the interface

interface, it follows that

$$-\frac{\partial}{\partial x}\mathrm{Re}I_m=\frac{\alpha_m\beta_m}{\mu_m\omega}E_0^2(1+\rho^2+2\rho\cos\varphi)e^{-2\alpha_m x} \tag{3}$$

Electromagnetic Diathermy

Dielectric studies have been carried out above 1,000 megacycles[6–8] and below 100 megacycles[9] for various body tissues. In addition, blood has been investigated[10] in the range 100 to 1,000 megacycles. The mechanism of absorption of electromagnetic energy is well enough understood[11] to be able to state that all tissues with high water content will have properties similar to that of blood in the 100 to 1,000-megacycle range. To complete the picture for fatty tissue, it was necessary to measure the dielectric constant and resistivity of fat in the range 100 to 1,000 megacycles. The methods of measurement will be reported elsewhere. The relative dielectric constant and resistivity in ohm-centimeters for horse and pork fat and muscle are summarized in Tables I and II. Although muscle impedance is remarkably uniform, the range of values for various kinds of fat is of the order of a factor of two. All measurements were obtained at temperatures near 37 degrees centigrade with fresh tissue samples. The data for the water-rich horse fat are used for the numerical calculations which follow.

The propagation constant can be written in terms of these data as

$$\gamma=j(2\pi/\lambda)\sqrt{\epsilon^+}$$

where

$\epsilon^+=\epsilon'-j\epsilon''$ is the complex dielectric constant
ϵ' = the relative dielectric constant
$\epsilon''=60\lambda_0\kappa$
λ_0 = the electromagnetic wave length in air
κ = the electrical conductivity

Values of the attenuation constant α and the phase constant β for horse fat and muscle are given in Fig. 2. From these values it is possible, using equations 2 and 3, to predict the heat development as a function of depth of tissue for a fat-muscle configuration such as shown in Fig. 1.

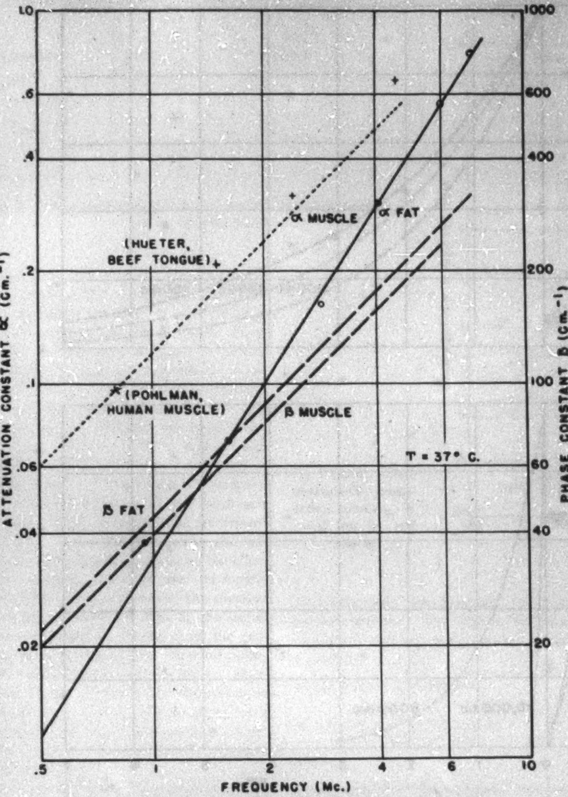


Fig. 4. Acoustic propagation constant $\delta = \alpha + j\beta$ for fat and muscle

The results for frequencies 150 megacycles to 10,000 megacycles are presented in Fig. 3. All values are plotted relative to the heat development in the muscle at the fat-muscle interface. In the electromagnetic case, the reflection factor is high. The magnitude ρ varies from 0.5 to 0.6 throughout the frequency range. This gives rise to a marked standing wave pattern which is predominantly important in determining the heating pattern in the fatty layer. The phase angle is roughly 180 degrees, which tends to minimize heat development in fat near the boundary. It is apparent that the lower frequencies are indicated for a high depth of penetration of energy into the muscle and low heating of the fatty layer. The advantage of radiation diathermy is illustrated by the fact that even at 3,000 megacycles, where the depth of penetration in muscle is very small, it is still possible to get some energy into the first centimeter of the muscle layer even though it is covered by as much as 2 centimeters of fat. The curves, however, show that lower frequencies are much more effective in providing deep heating.

Ultrasonic Diathermy

In the acoustic case, free-field methods are ordinarily used to determine the propagation constant directly. Comparatively little information is available in literature regarding the ultrasonic properties of tissue. Hueter[12] has shown that the absorption of muscle tissues is a linear function of frequency. Pohlman[13] gives a value for the absorption coefficient of human muscle at 800 kc. For the absorption constant of muscle α_m in Fig. 4, a linear curve has been drawn through Pohlman's 800-kc value. Also shown are Hueter's values for beef tongue. Other muscle tissues which he measured have higher absorption coefficients. Ludwig[14] gives the value 1.68×10^5 centimeters per second for velocity of sound in beef muscle.

Essentially no information is available in the literature on the acoustic properties of fatty tissue. To carry out this study, measurements of the absorption and velocity of sound were conducted on fresh pork fat. The corresponding propagation constants for temperatures of 37 degrees centigrade are given in Fig. 4. It will be noted that the absorption in fat has approximately a 1.5 dependence upon frequency in the range from 1 to 7 megacycles. Since depths of penetration of the order of 4 centimeters are desirable for deep heating the frequency range near 1 megacycle is indicated for ultrasonic diathermy.

The density of the fatty tissue used in the measurements was approximately 0.93. Ludwig[14] gives a value of 1.06 for the density of muscle tissue. From these data and the propagation constants it follows, by equation 1, that only roughly 2 per cent of the incident energy is reflected at a fat-muscle interface in the acoustic case. Furthermore, the wavelength is of the order of 1 millimeter; hence any spatial fluctuations of heat development caused by standing waves would be so close together that they would be smoothed out by heat conduction. Using the values given in Fig. 4, the heat development for a fat-muscle configuration is computed by equations 2 and 3, and presented in Fig. 5, for frequencies of 1 megacycle and 2 megacycles.

Conclusions

Perhaps more important than heat development per unit volume is the total rate of heat development in the fatty layer, as compared with that in the muscle. This information is obtained simply by integrating equations 2 and 3. With thickness of the fatty layer as a parameter, the ratio of heat developed in a layer of fat to that in a semi-infinite region of muscle is shown for the electromagnetic case in Fig. 6 and for the ultrasonic case in Fig. 7.

If it is arbitrarily assumed desirable to generate at least three times as much heat in the muscle as in the fat, then frequencies of roughly 500 megacycles in the

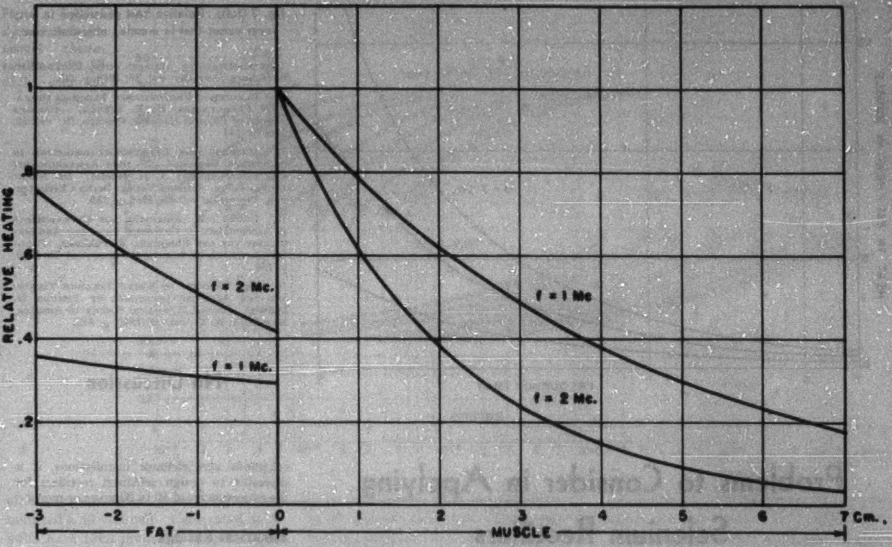


Fig. 5. Heat development per unit volume in the ultrasonic case. (All values are relative to heat development in muscle at the interface)

electromagnetic case and 1 megacycle in the ultrasonic case should be used. In short-wave diathermy the ratio of heat developed in fat to that in muscle decreases with increasing frequency. Just the opposite is true for radiation diathermy, where lower frequencies are recommended for selective heating of muscle. With short-wave diathermy, the fatty layer in a series fat-muscle configuration is always selectively heated. However, with both forms of radiation diathermy it is possible, by proper choice of frequency, to penetrate the fatty layer and dissipate a large fraction of the total energy in the underlying muscle.

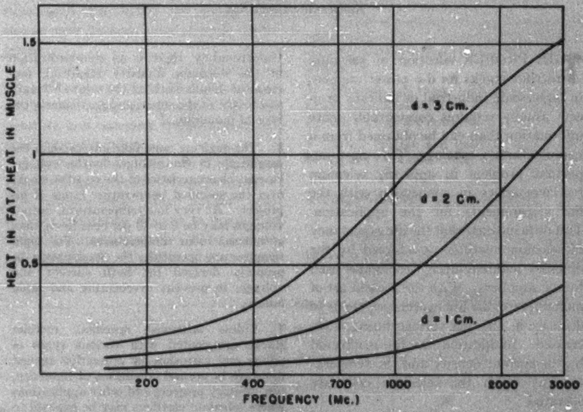


Fig. 6. Relative heat generation in fatty layer versus that in muscle; electromagnetic case

References

1. Temperaturmessungen an geschichteten biologischen Geweben bei Frequenzen von 2.7×10^7 cps bis 1.2×10^8 cps, **A. Esau, J. Paetzold, E. Ahrens.** *Die Naturwissenschaften*, Springer-Verlag, Berlin-Charlottenburg, Germany, vol. 24, 1936, p. 520.

2. Application of Electric and Acoustic Impedance Measuring Techniques to Problems in Diathermy, **Herman P. Schwan, Edwin L. Carstensen.** *AIEE Transactions*, vol. 72, part I, May 1953, pp. 106–10.

3. Transmission Calculus and Electrical Transients (book), **S. Goldman.** Prentice-Hall, Inc., New York, N. Y., 1949, p. 411.

4. Electromagnetic Theory (book), **J. H. Stratton.** McGraw-Hill Book Company, Inc., New York, N. Y., 1941.

5. Fundamentals of Acoustics (book), **A. R. Frey.** John Wiley and Sons, Inc., New York, N. Y., 1950, p. 141.

6. Dielectric Properties of Tissues Important in Microwave Diathermy, **J. F. Herrick, D. G. Jelatis, G. M. Lee.** *Federation Proceedings*, The Federation of American Societies for Experimental Biology, Washington, D. C., vol. 9, 1950, p. 60.

7. Dielectric Properties of Human Body for Wavelengths in the 1–10 cm. Range, **T. S. England.** *Nature*, London, England, vol. 166, 1950, p. 480.

8. The Dielectric Behavior of Same Types of Human Tissue at Microwave Frequencies, **H. F. Cook.** *British Journal of Applied Physics*, London, England, vol. 2, 1951, p. 295.

9. High Frequency Conductivity and Dielectric Constant of Biological Tissues and Fluids, **K. Osswald.** *Hochfrequenztechnik und Elektroakustik*, Leipzig, Germany, vol. 49, 1937, p. 40.

10. Die Dielektrizitaetskonstante und Leitfaehigkeit des Blutes bei ultrahohen Frequenzen, **B. Rajewsky, H. P. Schwan.** *Die*

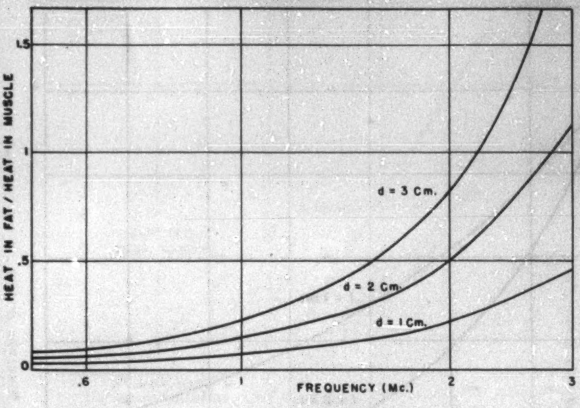


Fig. 7 (left). Relative heat generation in fatty layer versus that in muscle; ultrasonic case

Naturwissenschaften, Springer-Verlag, Berlin-Charlottenburg, Germany, vol. 35, 1948, p. 315.

11. Electrical Properties of Blood at Ultrahigh Frequencies, H. P. **Schwan.** *American Journal of Physical Medicine*, Chicago, Ill., vol. 32, 1953, p. 144.

12. Messung der Ultraschallabsorption in tierischen Geweben und ihre Abhaengigkeit von der Frequenz, T. H. **Hueter.** *Die Naturwissenschaften*, Springer-Verlag, Berlin-Charlottenburg, Germany, vol. 35, 1948, p. 285.

13. Ueber die Absorption des Ultraschalls in menschlischen Geweben und ihre Abhaengigkeit von der Frequenz, R. **Pohlman.** *Physikalische Zeitschrift*, Leipzig, Germany, vol. 40, 1939, p. 159.

14. The Velocity of Sound Through Tissues and the Acoustic Impedance of Tissues, G. **Ludwig.** *Journal*, Acoustical Society of America, New York, N. Y., vol. 22, 1950, p. 862.

◆

No Discussion

Problems to Consider in Applying Selenium Rectifiers

J. GRAMELS
ASSOCIATE MEMBER AIEE

THE PROPER selection of selenium rectifier stacks for d-c power supplies in large scale industrial or military projects usually requires considerably more information than can be obtained from a manufacturer's catalog. The most important problem in applying selenium rectifiers arises in connection with the life requirements for the application. Test data indicate that the life expectancy of selenium rectifiers is affected by the different manufacturing techniques used by the suppliers. With the present art of manufacture, the life expectancy tends to decrease as the cell voltage rating is increased. In addition, the life is affected by the current density and the temperature at which the selenium cells are operated.

The selection of selenium rectifier stacks for use in d-c power supplies involves several important considerations, as follows:

1. Circuit requirements must be carefully analyzed so that allowances may be made for variations in the voltage-current characteristics of each manufacturer's product, as well as for variations that exist for the same stacks processed by different manufacturers.

2. The equipment engineer must anticipate the differences in mechanical details of the same stack assembled by different suppliers. Unfortunately, there is no standardization in the selenium industry regarding mechanical details such as the over-all length and height of the stack, and particularly the type of mounting.

3. The engineer must take into account the magnitude of the changes in the voltage-current characteristics of the rectifier stacks over the specified temperature range of his project. At very low temperatures, output voltages may be 5 to 10 per cent lower than at normal room temperatures. For high-temperature operation, the stacks must be properly derated for both current and voltage, to prevent overheating and rapid failure.

4. Unless otherwise specified, rectifier stacks are coated with various types of paints and varnishes for protection against moisture in normal conditions of humidity. For military projects and other applications where selenium rectifiers may be exposed to high humidities, fungus, salt, or other corrosive atmospheres, the rectifier stacks must be provided with a more suitable type of protective coating or finish. These finishes are available from most manufacturers.

5. Selenium rectifiers age with time. Compensation for this aging should be provided if load requirements warrant. The project engineer should determine what life he expects or requires of the application. For military applications life requirements may vary from minutes to thousands of hours. On other applications, such as telephone and elevator installations, it is desirable to design selenium rectifiers for life expectancies of 10 to 20 years or more.

Electrical Ratings

The electrical ratings of selenium rectifiers are based on their voltage, current, and thermal characteristics. All three of these characteristics must be considered carefully for initial design purposes, as any one can affect life expectancy.

Voltage Ratings

The voltage rating is usually expressed as the reverse voltage rating. It is the maximum rms sine-wave voltage above which an excessive reverse current would flow and overheat the cell, causing breakdown. However, the important consideration establishing the rating is the peak voltage applied to the cell. If the applied voltage differs significantly from a sine-wave, it is important that the peak voltage shall not exceed 1.41 times the rated rms voltage.

When selenium cells were originally manufactured in this country, their rms voltage ratings ranged from 14 to 18 volts. Ratings later increased to 26 volts, and at the present time the goal appears to be 33 volts or higher. One manufacture has already successfully produced 33-volt cells for several years. An-

---

Paper 53-215, recommended by the AIEE Metallic Rectifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15-19, 1953. Manuscript submitted October 24, 1952; made available for printing April 14, 1953.

J. Gramels is with the Bell Telephone Laboratories, Inc., New York, N. Y.

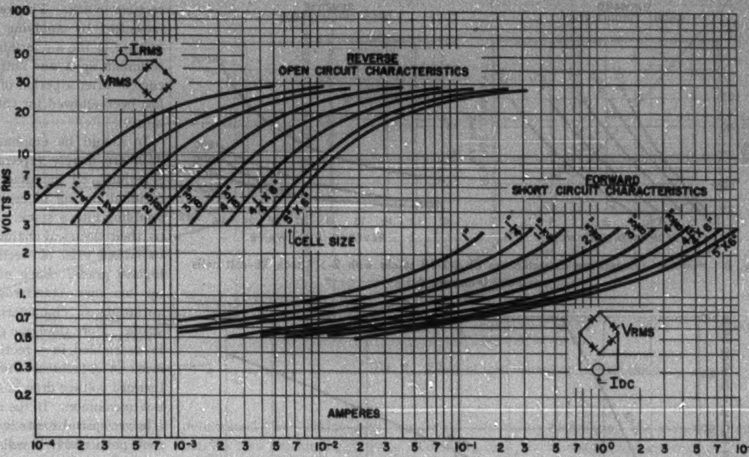


Fig. 1. Typical dynamic voltages current characteristics of 4-1-1B stacks with 26-volt cells

other manufacturer recently has announced a 40-volt cell. Cells rated at 70 volts have been produced in the laboratory. For uncritical applications, such as low cost radio and television sets, some manufacturers process cells rated at 45 volts rms. Generally, in such applications, a high quality produce cannot be expected and the units are likely to have a relatively short life.

Current Ratings

The current rating of selenium cells is based upon a normal current density of 0.32 ampere per square inch of active rectifying surface for a single-phase full-wave rectifier operating into a resistance load. Most manufacturers produce sizes varying from round 1/4-inch-diameter cells to 5- by 6-inch rectangular cells. Larger cell sizes are available from some manufacturers.

Thermal Ratings

The selenium rectifier is a thermal, as well as an electrically rated device. The voltage and current ratings are usually based upon operation of the rectifier stack in a normal ambient temperature of +35 degrees centigrade. It is important to note, however, that for selenium cell applications, ambient temperature is defined as the temperature immediately surrounding the stack within the equipment enclosure, not the temperature area where equipment is installed. Above +35 degrees centigrade ambient, the voltage or current ratings, or both, must be reduced. Rectifier stacks operating at rated current and voltage into a resistance load have temperature rises ranging from 15 to 35 degrees centigrade. The temperature rise depends upon the manufacturing techniques used by the various companies as well as the spacing of the selenium cells on the assembled stack. For long life expectancy, the actual cell temperature should not exceed +75 degrees centigrade.

Selenium rectifiers can be operated at cell temperatures above 75 degrees centigrade, but aging accelerates with temperature and excessive temperatures will result in rapid failure.

A special problem today is the application of selenium rectifiers to high temperature military uses. Many military projects are requesting designs to operate at ambient temperatures of +70 to +90 degrees centigrade. Unfortunately, reliable information about rectifier operation and life expectancy under these conditions is rare and generally not available. Statistical data should be accumulated and development studies should be undertaken to determine:

1. The life expectancy of selenium rectifiers when operated at the manufacturer's rating in ambient temperatures up to +100 degrees centigrade.
2. Whether, at these high temperatures, longer life can be obtained if the voltage and current ratings are reduced.
3. What voltage and current ratings can be applied to obtain longer life.

Voltage-Current Characteristics

To demonstrate the nonlinear characteristics of selenium rectifier cells, Fig. 1

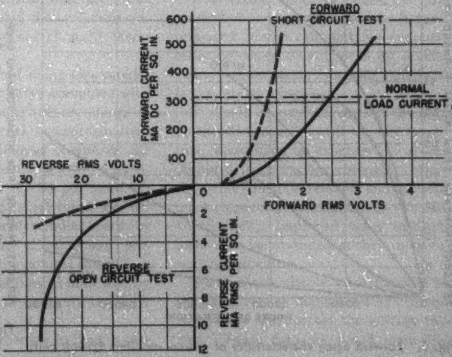


Fig. 2. Typical dynamic voltage-current characteristics of 4-1-1B stacks with 26-volt cells processed by two different suppliers

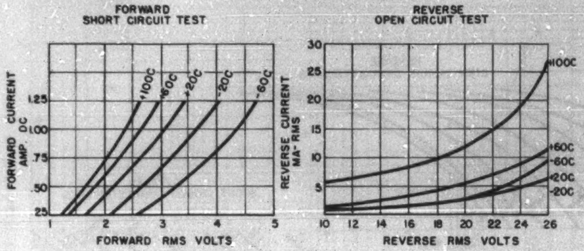


Fig. 3. Ambient temperature characteristics of 4-1-1B stacks with 2-×2-inch 26-volt cells rated at 1.0-ampere d-c

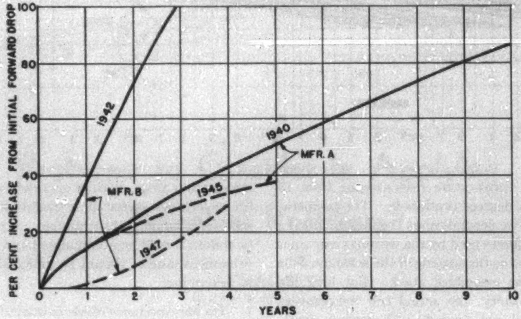


Fig. 4. Forward aging characteristics cell rating, 18 volts rms

illustrates representative dynamic voltage-current characteristics taken at room temperature for typical cell sizes rated at 26 volts, rms.

These data and subsequent results are plotted on the basis of unit type *4-1-1B* rectifier stacks. These stacks are single-phase full-wave rectifier circuits with a single selenium cell in each of the four rectifying elements. Tests were made with 60 cycle sinusoidal supply voltage. The forward voltage drop is expressed as the rms volts required to produce a specified current in a moving coil d-c ammeter connected directly to, and short-circuiting the output terminals of the rectifier stack. The reverse current is measured by applying a specified rms voltage to the a-c terminals with the d-c terminals open, and noting the rms input current after a period of 5 to 10 minutes has been allowed for the current to stabilize.

It should be emphasized that these data are merely typical characteristics. Selenium cell manufacture requires individual testing of each cell before assembling into a stack. Cells are graded by their electrical characteristics. Large variations exist between the lowest and highest grade. Each manufacturer sets up his own standards regarding the variations of characteristics in a given grade. More than one grade of cell may be assembled in a rectifier stack. Differences as great as ±50 per cent in the forward voltage drop at rated current are not uncommon. In the reverse direction, a larger spread exists in the reverse current, particularly in cells or stacks made by different suppliers.

Fig. 2 shows dynamic characteristics plotted on a linear scale to illustrate variations of the forward voltage and reverse current characteristics of selenium rectifier stacks that are processed by two different suppliers. Variations of this magnitude exist, not only from supplier to supplier, but also in a particular supplier's rectifiers. For a fixed a-c input voltage to these rectifier stacks operating into a resistance load, these differences in the forward voltage drop can vary the d-c output voltage ±5 per cent from the mean value. If this cannot be tolerated, selenium cells have to be carefully graded and selected to obtain uniformity, or other circuit adjustments have to be provided. Special selection of cells, obviously, will increase the cost of the product.

Selenium rectifiers, unlike most metal conductors, have a negative temperature

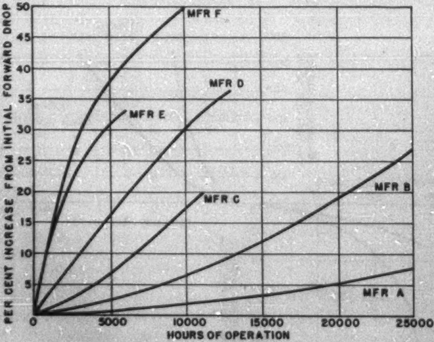


Fig. 5. Forward aging characteristics of various suppliers 26-volt cells

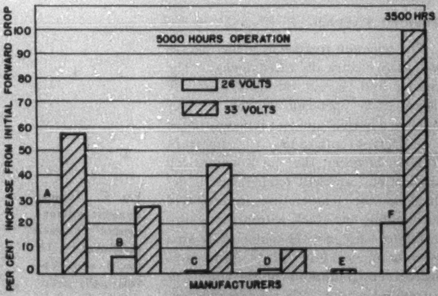


Fig. 6. Forward aging characteristics comparison of 26- and 33-volt cell ratings

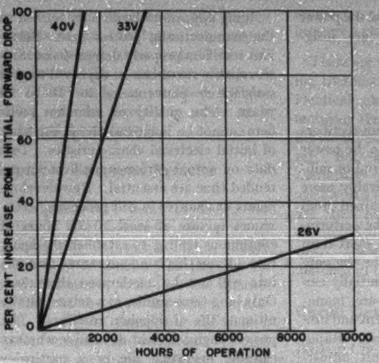


Fig. 7. Forward aging versus cell voltage rating

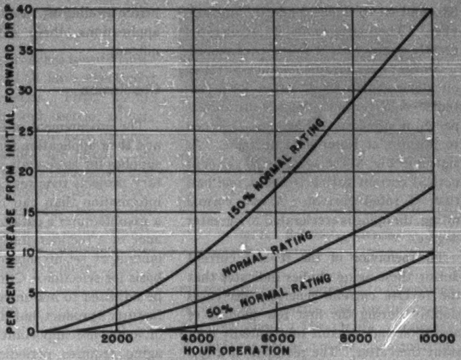


Fig. 8. Effect of load current on forward aging

coefficient of resistance in the forward direction. The forward voltage drop at a specified current decreases as the ambient temperature increases, shown in Fig. 3. In the reverse direction, the reverse current decreases as the temperature is lowered to approximately −20 degrees centigrade. Below this temperature, there is no apparent change in the current except at the higher voltages. At the higher voltages, the current again tends to increase.

Aging

Selenium rectifiers are subject to aging. The important factor in selenium rectifier aging is the increase in the forward voltage drop which results in a decreased d-c output. For normal rectifier applications, aging of the reverse current is not critical. If the rectifier circuit conditions require that the output voltage must be maintained over a period of time, provisions must be made to supply additional a-c input voltage to the rectifier stack as it ages. One or more taps on the associated transfer can compensate for this change. For example, on a rectifier stack operating at its rated d-c output current and voltage, it is considered good practice to design for a 100 per cent increase in the initial forward voltage drop of the stack. This requires approximately a 10 per cent increase in the a-c input voltage to maintain the initial d-c output. At this point, the a-c input voltage usually reaches the maximum reverse voltage rating. A further increase in the a-c voltage may result in overheating and rapid failure.

The extent and rate of aging of selenium rectifiers cannot be determined mathematically. Aging characteristics must be determined by actual test involving lengthy time-consuming projects. To determine whether a given rectifier stack will give satisfactory performance for 5 years, as an example, tests would have to be conducted for this period. Aging can be accelerated by operation at high temperatures or at load currents above normal, but no accepted correlation exists between this type of aging and that obtainable under normal operating conditions.

The following aging data were obtained on sample rectifier stacks obtained from different manufacturers in this country. The samples were set up on life test racks as single-phase full-wave rectifiers and were operated continuously at normal room ambient temperatures. For the duration of the tests, the a-c input voltage to the stacks was kept constant at approximately 10 per cent below the maximum rms voltage rating. The stacks operated into a resistance load. The resistance was selected so that the rectifiers operated at rated d-c load currents. The resistance was not changed thereafter.

Long-term forward aging characteristics of selenium rectifiers assembled with cells rated at 18 volts rms are shown in Fig. 4. After 10 years, or approximately 90,000 hours, continuous operation, the forward drop on one supplier's product increased approximately 90 per cent. A second supplier's product increased 100 per cent in 3 years under similar operating conditions. Further development programs by each manufacturer resulted in improved aging qualities as indicated by tests made on stacks obtained at a later date.

By introducing changes in their manufacturing techniques, the industry eventually increased the voltage ratings to 26 volts rms per cell. This change in processing, however, raises the question, how has the operating life been affected? Fig. 5 clearly illustrates that although many manufacturers are commercially producing 26-volt cells, the life expectancy is vastly different from the six suppliers' stacks that were investigated.

In order to save space, weight, and critical materials, manufacturers are attempting to produce cells with still higher voltage ratings. Obviously, the higher the voltage rating, the fewer cells are required in series for a given a-c input voltage. Most companies have had a 33-volt cell development program under way for some time, but again the problem of aging must be considered. Fig. 6 compares the forward aging of 26- and 33-volt cells after 5,000 hours of operation. It is evident that in all cases except one, the 33-volt cells age substantially faster than the 26-volt cells. The exception is the product of one manufacturer who has not produced 26-volt cells but has commercially processed 33-volt cells for several years. Again, there is a wide difference in the rate of aging.

It should be emphasized that the data on the 33-volt cells, with the one exception, were taken on experimental stacks obtained from various companies while they were still in the research and development stages of processing the 33-volt cell. Undoubtedly, improved aging qualities have been incorporated in their more recent manufacture.

Fig. 7 shows clearly what happens to the aging rate as manufacturers attempt to increase still further the voltage ratings. These data are based upon 26-, 33-, and 40-volt cells made by one supplier. The aging rate is accelerated greatly as the voltage rating is increased.

All the foregoing aging data were taken on stacks operated at rated load currents. A longer operating life may be expected if selenium rectifiers are designed to operate at lower load currents. As shown in Fig. 8, the aging at 50 per cent normal current rating is about one half that at rated current. Above normal rating, the aging is accelerated to a greater degree.

The behavior of the reverse current during these aging studies indicates that the reverse current generally decreases slightly during the first few months of operation and remains substantially constant thereafter. The rectifiers of a few manufacturers, however, show the opposite effect. The current rises slightly before leveling off. For normal d-c power applications, these changes are insignificant.

Conclusion

The procurement of selenium rectifiers and their application for use in d-c power supplies for large-scale industrial or military projects requires considerably more information than can be obtained from a manufacturer's catalog. The performance of a rectifier stack in a short-term laboratory set-up should not be the only basis for selection. Circuits initially can be adjusted to accommodate any manufacturer's product, but the extent and rate of aging are important factors. Rapid aging requires periodic circuit adjustments and more frequent replacement of the stacks.

Aging appears to be directly related to the manufacturing process used. Selenium rectifiers properly designed and conservatively rated can be expected to give satisfactory performance for 10 to 20 years. The quality of selenium rectifiers cannot be judged solely on the basis of initial electrical characteristics. Test data on actual performance over an extended time are essential. New developments or changes in cell processing techniques require at least 10,000 hours of continuous testing to establish the aging rate. Even this is no guarantee that the rate will not be accelerated thereafter. Only long-term studies can determine the ultimate life of selenium rectifiers. The project engineer must determine whether suitable operating life of a particular manufacturer's product will meet the requirements of his application.

No Discussion

40- to 4,000-Microwatt Power Meter

R. W. LANGE
NONMEMBER AIEE

Synopsis: This paper describes a 10-kilomegacycle power meter using a bead thermistor for a power detector and having full scale sensitivities from +6 decibels referred to 1 milliwatt (dbm) (4,000 microwatts) to −14 dbm (40 microwatts) selected in 2-decibel (db) steps. Absolute calibration from built-in d-c standards is provided.

Ease of operation at high sensitivities is accomplished by the improvement in zero stability resulting from: 1. Better temperature compensation by closer proximity of temperature and power measuring thermistors; 2. Selection of thermistor pairs after aging; 3. Adjustment of circuit parameters to compensate for manufacturing variations in thermistor parameters; and finally 4. Deriving entire thermistor bias power from a stabilized bridge oscillator.

MICROWAVE power measurements have been described in the technical literature for measuring powers from kilowatts to a fraction of a milliwatt. High powers may be measured conveniently by calorimetric methods; low powers by bolometric methods.

The power meter to be described is of the bolometric type. It extends the accuracy and convenience of measuring in the milliwatt range into the microwatt range.

Bolometer Bridge Circuit

A simplified block diagram is shown in Fig. 1. This is a temperature-compensating direct-reading bridge using a self-balancing oscillator. A brief review of the operation of this circuit follows.

The bolometer detector is a thermal resistor known as a bead thermistor, matched to terminate a transmission line with low reflection. It is shown in this figure as the element marked *RF*. It is the only element in the *RF* resistance bridge subjected to the microwave power.

Directly below the ground plane is the compensating bridge, or *COMP* bridge. Idealized, this bridge is the mirror image of the *RF* bridge except that the *COMP* thermistor is not subjected to the microwave power. The two bridges connected in series provide the load on the amplifier. The output of the *RF* bridge is connected in the proper phase to the input of the amplifier. The thermal resistor in the bolometer detector is used as the regulating element in a conventional bridge-regulated oscillator, holding the resistance of the thermistor very close to the bridge-balancing value. The oscillator power varies with the ambient temperature of the thermistor. Consider the operation before application of microwave power. In the idealized case where the two thermistors have identical characteristics and operate at the same ambient temperature, the identical bridges are simultaneously balanced.

With the application of microwave power to the *RF* thermistor, the *RF* bridge is automatically balanced as before by a reduction in the oscillator power equal to the applied microwave power. This same reduction in oscillator power is applied to the *COMP* bridge and, since no microwave power is applied

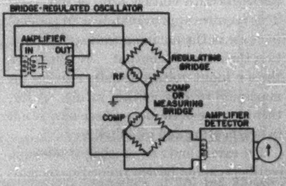


Fig. 1. Bolometer bridge circuit

Paper 53-281, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted February 18, 1953; made available for printing April 29, 1953.

R. W. LANGE is with the Bell Telephone Laboratories, Inc., Murray Hill, N. J.

The author acknowledges the pioneering design of temperature-compensating bolometer bridge circuits and of the basic principles of this power meter by E. W. Houghton and his associates, also the work on stabilization of bead thermistors by C. B. Green and his associates, all of Bell Telephone Laboratories.

to the *COMP* thermistor, this bridge operates at an unbalance proportional to the microwave power applied to the *RF* thermistor. The unbalance voltage of the *COMP* bridge is amplified and detected and read directly as power on a calibrated d-c meter.

Improved Bolometer Detector

The microwave heart of the power meter is the bolometer detector. A cutaway view is shown in Fig. 2. There is the need to operate the thermistors, even thermistors with identical characteristics, at the same ambient temperature. A typical thermistor will have a power temperature coefficient of 100 microwatts per degree centigrade. If we wish to measure 40 microwatts, it is desirable to hold the error in the power indication arising from the random ambient temperature difference of the *RF* and *COMP* thermistors to 2 microwatts or less. This corresponds to an ambient temperature tracking of 0.02 degree centigrade.

In the detector of Fig. 2 the thermistor beads are nearly superimposed. This places the beads at essentially the same ambient temperature. The leads to the beads are similar to give like heat conduction, thereby better ambient temperature tracking. Any residual difference in the thermal conductances of the thermistors to the surroundings appears as a difference in the power-temperature slopes and is corrected, as described later, by the circuit parameter R_∞. The *RF* thermistor is coupled in the usual manner to the TE_{10} mode of the wave guide. The *COMP* thermistor is not coupled to the TE_{10} mode and therefore does not absorb microwave energy. This design has contributed materially to improve the ambient temperature tracking of the beads, making the 40-microwatt sensitivity practical.

General Thermistor Characteristics

The next practical consideration is that thermistor beads do not have identical characteristics. Some of the electrical characteristics are:

1. Aging.
2. The absolute bias power needed to operate the thermistor at a particular resistance for a stated temperature.
3. The power-temperature characteristic which is the relationship between bias power and temperature with resistance a parameter.
4. The resistance-power characteristic which is the relationship between resistance change per unit applied power with temperature a parameter.

There are two approaches toward causing thermistor beads to have effectively identical characteristics. First, manufacturing them alike and, second, adjusting circuit parameters to fit each thermistor. Both approaches are used. The thermistor beads are manufactured and aged with a high degree of control. They are then selected as pairs having like characteristics and similar aging tendencies. These pairs are mounted and matched in the microwave power head.

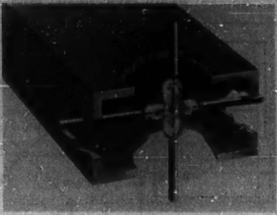


Fig. 2. Bolometer detector

Circuit Description

Bridge Circuit

For the discussion on adjustments of the circuit parameters, refer to Fig. 3 which shows the circuit in greater detail. The amplifier section of the bridge-regulated oscillator must provide stable gain and deliver the total power required to operate the two bridges. The amplifier is stabilized by negative feedback and is operated from a regulated power supply. Both thermistors are biased entirely by the oscillator power.

Circuit Parameters for Accommodating Thermistor

The *RF* thermistor is held at the resistance required to match it into the bolometer detector. A typical matching resistance value is 125 ohms. The *COMP* thermistor is held at a resistance determined by R_c. This value is derived as follows: In Fig. 4 the power-temperature characteristics of a pair of thermistors are shown. At temperature t_0 no bias power is required to operate thermistor *RF* at 125 ohms; t_0 is around 300 degrees Fahrenheit, well above the maximum ambient operating temperature. At lower temperatures bias power must be added to maintain the resistance at 125 ohms since at t_0 no heating power is applied. For this particular thermistor operating at 125 ohms the thermistor bead must be held constant at temperature t_0, and this temperature is attained by the combination of ambient temperature and total bias power. The *COMP* thermistor, which is the second one of the pair, requires a temperature t_1 to reach a resistance of 125 ohms. Since the two thermistor beads are at the same ambient temperature and have like heat conduction properties, they can be made to temperature-track over a large operating range by shifting the compensating thermistor characteristic to temperature t_0. To do this the *COMP* thermistor operating resistance corresponding to t_0 becomes higher than 125 ohms, say, for this example, 140 ohms. The circuit parameter R_c, which is one arm of the compensating bridge, is made 140 ohms for this particular thermistor, thereby balancing this bridge when the *COMP* thermistor is 140 ohms. R_c is individually determined for each thermistor pair.

Next, observe that the slopes of the two curves are different. This means that for any increment in temperature the *COMP* thermistor requires a greater change in bias power from the oscillator to maintain its resistance than does the *RF* thermistor. The circuit parameter R_∞ provides that the ratio of the two required powers is the same as the ratio of the slopes of the power-temperature characteristics. This division of oscillator power is made by the variable shunt around each bridge. R_∞ is adjustable for each thermistor pair.

Amplifier-Detector

Three of the four characteristics of thermistors have been accounted for. In considering the fourth, recall in the review of this general type of power meter that the *COMP* bridge operates at an unbalance proportional to the microwave power applied to the *RF* thermistor. This unbalance is dependent on the resistance-power coefficient of the thermistors. To use practical thermistors the circuit must provide an absolute power calibration and adjustments of the amplifier detector sensitivity. Since the thermistor is unaware of its source of heat, the absolute power is conveniently supplied from the d-c calibrating circuit. This circuit consists of a selected network of precision resistors to derive calibrating powers in the range 40 to 4,000 microwatts from the d-c power supply. The sensitivity of the amplifier-detector is adjusted by varying the cathode feedback in the preamplifier, or the zero-db adjust control, which is adjusted for each thermistor pair.

The amplifier-detector is a 3-stage

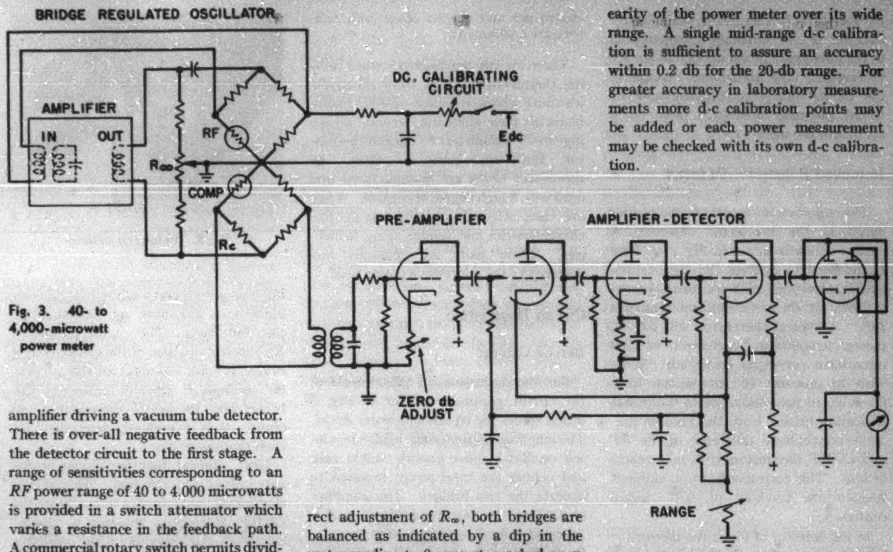


Fig. 3. 40- to 4,000-microwatt power meter

amplifier driving a vacuum tube detector. There is over-all negative feedback from the detector circuit to the first stage. A range of sensitivities corresponding to an *RF* power range of 40 to 4,000 microwatts is provided in a switch attenuator which varies a resistance in the feedback path. A commercial rotary switch permits dividing this range into 2-db steps from −14 to +6 dbm. This range switch is shown functionally in Fig. 3 as the variable resistor designated *RANGE*.

Operating Procedure

There are three simple calibrating steps before applying and measuring the microwave power. First, the supply voltage is adjusted to the calibrate value as marked at full scale on the meter. Second, with the calibrate switch in the adjust ∞ position, R_∞ is adjusted for a meter indication of ∞. (With the correct adjustment of R_∞, both bridges are balanced as indicated by a dip in the meter reading to 0 current, marked as ∞ on the meter scale.) Third, the calibrate switch is operated to apply a known d-c power to the *RF* thermistor. The zero-db adjust control is adjusted to give a full-scale meter indication of 0 db. The actual full-scale power sensitivity is indicated on the range switch. This completes the calibration.

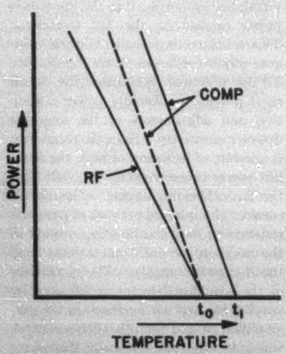


Fig. 4. Power-temperature thermistor characteristics

For measurement, the calibrate switch is operated to the measure position. The microwave power under test is applied to the *RF* thermistor. The sum of the meter indication and the setting of the range switch gives the measured microwave power in decibels referred to 1 milliwatt. The meter scale is calibrated in 0.1-db intervals to 2 db and 0.5-db intervals to 6 db, so power ratios 6 db below 14 dbm, that is −20 dbm, or 10 microwatts, may be read directly.

Stability

The stability of the power meter after an initial warm-up of less than 2 hours is such that for the next 4 hours the drift on the most sensitive scale of −14 dbm is of the order of 1 microwatt. The stability is of the same order with 3 per cent changes in filament and plate voltage. However, these voltages can be regulated quite readily to cause negligible drift.

Linearity

Selection of impedance of the bridge circuits and use of feedback around the amplifier detector contribute to the linearity of the power meter over its wide range. A single mid-range d-c calibration is sufficient to assure an accuracy within 0.2 db for the 20-db range. For greater accuracy in laboratory measurements more d-c calibration points may be added or each power measurement may be checked with its own d-c calibration.

Performance on Continuous-Wave and Pulse Power

Power measurements on continuous-wave signals have been checked against a thermistor detector operating in a simple d-c bridge circuit where the microwave power is the difference in d-c bias power delivered to the thermistor with and without microwave power applied. The two power meters were in agreement within 0.05 db.

Power measurements were made using pulse signals from magnetrons. Pulse lengths of 0.25 to 0.5 microsecond and repetition rates from 500 to 2,000 cycles were used. Power measurements were compared with the d-c thermistor bridge and also with a water load calorimeter. The correlation of the three power measurements was within 0.25 db.

Conclusion

This direct-reading power meter has proved to be a stable, sensitive tool for both laboratory and field use. The basic circuit is suitable for operation with an extension of the design of the improved bolometer detector to other frequency bands. These same considerations apply to using hot-wire, or barretter, thermal resistors.

◆

No Discussion

Mathematical Descriptions of Core Losses

J. W. HALE
NONMEMBER AIEE

F. R. RICHARDSON
NONMEMBER AIEE

Synopsis: This paper presents results so far achieved by the writers in endeavoring to gain an understanding of the factors controlling a-c core losses in modern magnetic materials. Losses have been measured on a number of cores over a range of frequencies and at flux densities in the middle portions of the magnetization curves. Theoretical and empirical relations have been derived to connect the dynamic losses to the static losses and other determining factors.

It has been found that in those cases where the combined effect of gauge, frequency, flux density, permeability, and conductivity is sufficiently severe the skin effect during the major part of each cycle reaches a maximum value which is limited by practical saturation of the material near the surfaces of the laminations. A theoretical equation describing eddy currents accompanying this limiting flux distribution gives good agreement with losses measured under these conditions. This equation is analogous to one offered previously by Ganz.[1]

In cases where this limiting flux distribution does not prevail, eddy losses are of greater practical importance, and are more complicated. Data for these cases have been studied by means of dimensional analysis. By dividing the materials tested into groups with similarly shaped magnetization curves, significant correlations have been obtained between parameters embodying the classical variables which are known to influence core losses. It has been found that there also exists in each group a previously unidentified material variable of major importance, and it is suggested that this is domain size. In making this interpretation, the writers rely on the explanation of the eddy current anomaly of Williams, Shockley, and Kittel.[3]

Data are presented which indicate that core losses in modern magnetic materials would be reduced 40 per cent or more if the static hysteresis losses could be kept the same while reducing the domain size to that which exists in ingot iron.

IT HAS been known for many years that observed losses in ferromagnetic laminations are greater than the losses calculated by standard methods.[2] This discrepancy has been referred to as the eddy current anomaly. In the case of 0.014-inch oriented silicon steel at 60 cycles per second, it amounts to about 40 per cent of the observed losses. Stewart[2] has found that refinements in the standard calculations to account for nonuniform permeability and skin effect have not succeeded in explaining the discrepancy. More recently, Williams, Shockley, and Kittel[3] have made a study of losses in a single crystal of silicon steel, and have shown that there is no anomaly in this special case where eddy currents can be accurately calculated on the basis of the actual domain structure. They suggest that classical eddy current equations fail chiefly because flux changes in laminations are treated as being continuous instead of discretely located at moving domain walls.

Because standard calculations fail to describe core losses in modern magnetic materials, no mathematical relations have been presented which quantitatively evaluate the effects on core losses of material variables such as gauge, resistivity, or permeability. It was to gain such quantitative understanding of core losses and to secure knowledge of possible unrecognized material variables that this study was undertaken.

Empirical Relations Describing Core Losses

The method of dimensional analysis outlined by Buckingham[4] has been applied to core loss measurements to obtain empirical relationships describing core losses when flux distribution in laminations is not limited by saturation. In this approach, three independent dimensionless parameters have been derived from core losses and the six classical physical factors known to influence them. Core loss data collected on a number of magnetic materials over a range of frequencies have been used to evaluate these parameters. Two advantages are gained from this method of attack:

1. The presence or absence of a correlation between these three parameters indicates respectively the absence or presence of additional important factors which influence core losses.
2. Since one of the three parameters is constant within certain groups of materials, a 2-dimensional plot of the other two parameters shows the effects on core losses of all the independent classical variables within limits where a correlation exists.

The seven classical variables which would be contained in a complete physical equation describing core losses are as follows:

P_c = core losses, watts per cubic centimeter
f = frequency, cycles per second
B_0 = operating flux density, gausses
ρ = resistivity, ohm-centimeters
d = lamination thickness, centimeters
μ = static permeability, gausses per oersted
J_0 = area of static hysteresis loop, joules per cubic centimeter per cycle at B_0

Three independent dimensionless parameters which embody these variables are

$$\frac{P_c}{fJ_0}, \frac{fB_0^2d^2}{\rho J_0}, \frac{\bar{\mu}_0 J_0}{B_0^2}$$

Since μ is not a single-valued function, it is necessary to assign an average value to it. $\bar{\mu}_0$ is the permeability averaged with respect to flux density from 0 to B_0. B_0 is the average flux density that exists in a lamination at the instant of zero voltage.

Values of $(\bar{\mu}_0 J_0)/B_0^2$ for the materials tested are given in Table I. It can be seen there that the values of this parameter divide the materials into two distinct groups, the members of group [illegible]ll having values close to 1.0×10^{-8} and the members of group II having values close to 1.5×10^{-8}. The members of group II are all materials with rectangular hysteresis loops. The constancy of the parameter in either group is the result of the general similarity in shapes of magnetization curves and hysteresis loops that exists in each group. Within each group, the average permeability is so closely related to the hysteresis loss and flux density that it need not be treated as an independent variable. Therefore, within each group, a correlation between P_c/fJ_0 and $(fB_0^2d^2)/\rho J_0$ will show the effects on core losses of all the independent classical variables; and the absence of a correlation between these two parameters will demonstrate the presence of additional important material factors which influence core losses.

The range of flux densities over which the close relation between permeability and hysteresis exists is limited to the middle portions of magnetization curves, where magnetization proceeds by irreversible domain wall movements. For exam-

---

Paper 53-261, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 12, 1952; made available for printing April 17, 1953.

J. W. HALE is with the Allegheny Ludlum Steel Corporation, Brackenridge, Pa., and F. R. RICHARDSON is with the General Electric Company, Pittsfield, Mass.

The authors wish to thank D. E. Bovey for supplying original and corroborative core loss data used in preparing this report.

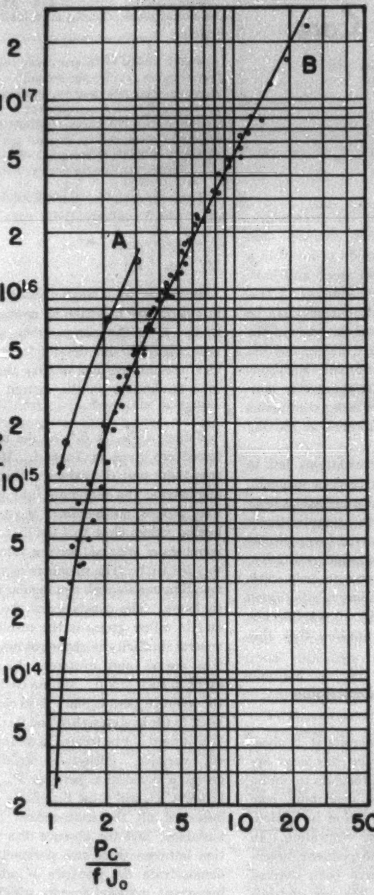


Fig. 1. Measured values of the dimensionless parameters for materials of group I. Curve A is from ingot iron data, and curve B from data for all other material of group I. The exact range of data for this and other figures is given in Table II

Table I. Values for Materials Tested of $\bar{\mu}_0 J_0/B_0^2$

Material	B_0	
	Group I	
0.002-inch unoriented Monimax	5,000	1.10×10^{-8}
0.002-inch 4–79 Moly-Permalloy	5,000	1.20×10^{-8}
0.004-inch unoriented Monimax	5,000	0.95×10^{-8}
0.005-inch oriented silicon steel	10,000	0.91×10^{-8}
0.0156-inch oriented silicon steel	5,000	0.98×10^{-8}
0.0156-inch oriented silicon steel	10,000	1.04×10^{-8}
0.010-inch Mumetal	5,000	1.00×10^{-8}
0.025-inch ingot iron	5,000	1.03×10^{-8}
0.025-inch ingot iron	10,000	1.11×10^{-8}
		1.04×10^{-8}
Average Mean Deviation = 7 per cent		
	Group II	
0.001-inch oriented Monimax	10,000	1.40×10^{-8}
0.002-inch oriented Monimax	5,000	1.55×10^{-8}
0.002-inch oriented Monimax	10,000	1.56×10^{-8}
0.002-inch Deltamax	5,000	1.43×10^{-8}
0.002-inch Deltamax	10,000	1.51×10^{-8}
0.003-inch oriented Monimax	5,000	1.53×10^{-8}
0.003-inch oriented Monimax	10,000	1.51×10^{-8}
0.004-inch oriented Monimax	5,000	1.52×10^{-8}
0.004-inch oriented Monimax	10,000	1.50×10^{-8}
0.005-inch Deltamax	5,000	1.51×10^{-8}
0.005-inch Deltamax	10,000	1.56×10^{-8}
0.002-inch Deltamax (magnetically annealed)	5,000	1.42×10^{-8}
0.010-inch Deltamax	10,000	1.46×10^{-8}
0.002-inch Deltamax (magnetically annealed)	10,000	1.63×10^{-8}
		1.51×10^{-8}
Average Mean Deviation = 3 per cent		

ple, this is from about 1,000 to 15,000 gausses in oriented silicon steel.

Empirical plots of P_c/fJ_0 and $(fB_0^2d^2)/\rho J_0$ are shown in Figs. 1, 2, and 3. Data for members of group I are given in Fig. 1; data for heavier gauge members of Group II are given in Fig. 2; and data for the lighter gauge members of group II are given in Fig. 3. The exact materials, flux densities, and frequencies covered in each plot are given in Table II. It will help in understanding these curves if it is noted that P_c/fJ_0 is the ratio of dynamic loss per cycle to static loss per cycle, and therefore must approach unity for small values of $(fB_0^2d^2)/\rho J_0$. Also, higher values of P_c/fJ_0 indicate higher relative losses.

In examining Fig. 1 it can be seen that a good correlation between the parameters exists for all the materials included there except ingot iron, and this clearly does not fit. It follows from the reasoning already presented that the significance of the correlation lies in its indication that no important variable besides those in the parameters acts to differentiate losses in the various materials included there. Conversely, the divergence of the ingot iron data demonstrates that some unknown material factor acts to differentiate its losses from those of the other materials in group I. Since only ingot iron losses are approximated by standard calculations, it is the authors' conclusion that this unknown material variable is responsible for the eddy current anomaly.

If the explanation of the eddy current anomaly[2] is accepted, then it can be reasoned that this unknown variable is domain size. For if the domains became very small, the conditions of the classical equation would be approached, and the losses would approximate those derived from the standard calculations. Also, the larger the domains, the greater the deviation from the assumed classical conditions, and the greater the eddy current anomaly. Photomicrographs of domain structures in 25-mil ingot iron and 12-mil oriented silicon steel, both taken at 200 diameters, are presented in Fig. 4 to demonstrate qualitatively that a large difference in domain size exists between these two materials.

From the principles of physical similarity described by Buckingham,[4] it follows that if domain size is a major variable influencing core losses, then the ratio of domain width to lamination thickness must remain constant to have a correlation between the parameters P_c/fJ_0 and $(fB_0^2d^2)/\rho J_0$. Therefore, the grouping of points shown in Fig. 1 suggests that those materials included there embody significant uniformity of effective domain size, and that this decreases proportionately with lamination thickness. The materials which show this correlation have in common, in addition to similarly shaped magnetization curves, the property of being single-phase alloys

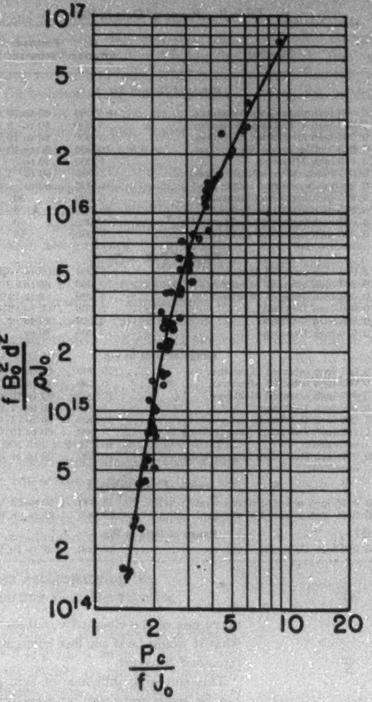


Fig. 2. Measured values of the dimensionless parameters for the 3-, 4-, 5-, and 10-mil materials of group II

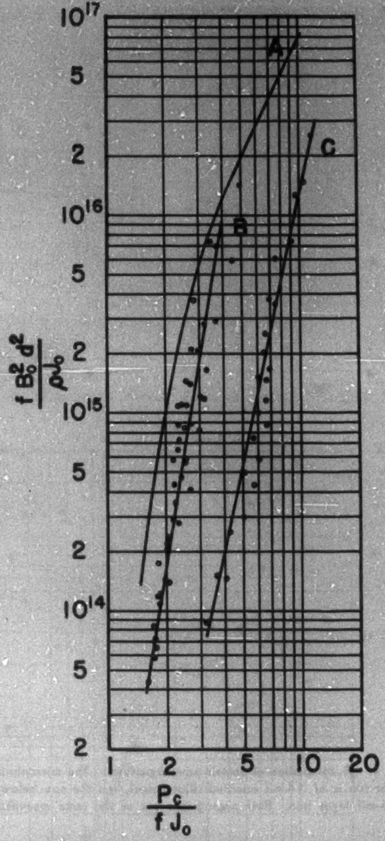


Fig. 3. Measured values of the parameters for the 1- and 2-mil materials of group II. Curve A is from Fig. 2. Curve B is for customarily annealed materials, and curve C for magnetically annealed Deltamax

undergoing no allotropic transformations, and have been high-temperature annealed.

In comparing the two curves in Fig. 1, it can be seen that the ratio of a-c to d-c loss per cycle for ingot iron is 40 per cent less than that for the other materials in group I. This indicates that if everything else could be kept the same in these metals while reducing the effective domain size to that which exists in ingot iron, then their losses would be reduced 40 per cent.

The effect of domain size also is a convenient and reasonable explanation for the displacements of loss curves which will be noted among the materials of group II plotted in Figs. 2 and 3. The heavier-gauge materials plotted by themselves in Fig. 2 form a good pattern, but the lighter-gauge materials in Fig. 3 form a parallel curve displaced slightly in the direction of higher losses. This could be a manifestation of a change in the ratio of domain width to lamination thickness, the domain width not decreasing as rapidly as the gauge. It fulfills a prediction that the eddy current anomaly may be greater in thinner sheets.[2]

The difference in losses of magnetically annealed Deltamax from that given ordinary furnace cooling, also shown in Fig. 3, is especially interesting, since it seems to be a case where core losses are increased by a factor of two through a treatment which presumably changes little, other than in the coarseness of the domain structure. Attempts by the authors to observe the domain structures in this material with powder patterns have not been successful.

If domain size can be accepted as being responsible for shifts in the parametric curves within each of the two groups of materials, then domain shape may be the factor differentiating losses between the two groups. This idea is strengthened by the fact that the rectangular hysteresis loops characteristic of the materials in group II are attributed to their special domain structures.

Although the empirical curves given in this paper have already been of considerable help in interpreting core loss phenomena, they constitute the beginning of a study, rather than the end. The relations may become valuable for estimat-

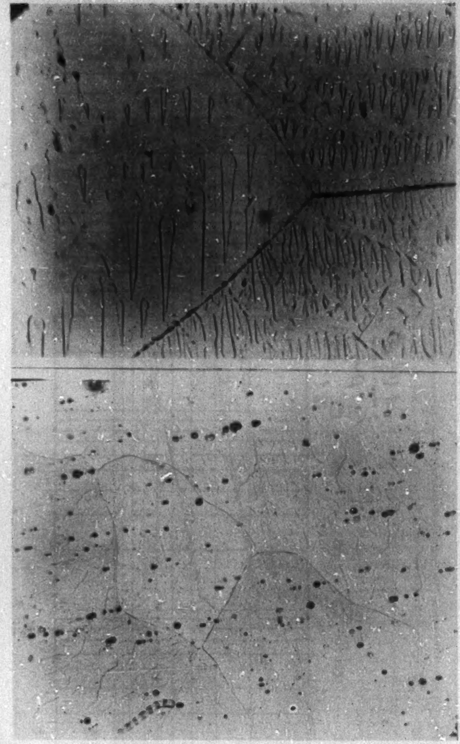


Fig. 4. A comparison of domain powder patterns. The microphotograph at the top is of 14-mil oriented silicon steel, and the one below is of 25-mil ingot iron. Both photographs are at the same magnification

Table II. Range of Data in Plots

Material	B_0, Gausses	Frequency Range, Cycles per Second
Range of Data in Fig. 1		
0.002-inch unoriented Monimax	5,000	60 to 70,000
0.002-inch 4–79 Moly-Permalloy	5,000	60 to 2,000
0.004-inch unoriented Monimax	5,000	60 to 10,000
0.005-inch oriented silicon steel	10,000	60 to 10,000
0.010-inch Mumetal	5,000	20 to 100
0.0156-inch oriented silicon steel	5,000	60 to 3,000
0.0156-inch oriented silicon steel	10,000	60 to 600
0.025-inch ingot iron (annealed)	5,000	60
0.025-inch ingot iron (annealed)	10,000	60
0.025-inch ingot iron (cold-rolled)	5,000	60
0.025-inch ingot iron (cold-rolled)	10,000	60
Range of Data in Fig. 2		
0.003-inch oriented Monimax	5,000	100 to 10,000
0.003-inch oriented Monimax	10,000	60 to 10,000
0.004-inch oriented Monimax	5,000	60 to 10,000
0.005-inch Deltamax	5,000	60 to 10,000
0.005-inch Deltamax	10,000	60 to 2,000
0.010-inch Deltamax	10,000	80 to 500
Range of Data in Fig. 3		
0.001-inch oriented Monimax	10,000	60 to 10,000
0.002-inch oriented Monimax	5,000	60 to 10,000
0.002-inch oriented Monimax	10,000	60 to 10,000
0.002-inch Deltamax	5,000	60 to 1,000
0.002-inch Deltamax	10,000	60 to 1,000
0.002-inch Deltamax (magnetically annealed)	5,000	60 to 10,000
0.002-inch Deltamax (magnetically annealed)	10,000	60 to 10,000
Range of Data in Fig. 5		
0.0156-inch oriented silicon steel	5,000	60 to 15,000
0.0156-inch oriented silicon steel	10,000	60 to 8,000
Range of Data in Fig. 6		
0.010-inch Deltamax	10,000	80 to 10,000

ing losses of untested materials when the limits of their use are more completely established. Accepting the importance of domain size and shape, the range of applicability of the relations will depend upon 1. the regularity within groups of magnetic materials of the ratio of domain width to lamination thickness, and 2. the number of significant kinds of domain configurations that exist in magnetic materials.

The Importance of Skin Effect

Skin effect in a magnetic lamination is the unequal magnetization through the thickness of the lamination which results from the fact that eddy currents shield central portions of the material from the full effect of the magnetizing force. It is always present under dynamic conditions of magnetization, but is small at very low frequencies, and becomes greater at higher frequencies. The actual patterns of flux distribution which occur during a cycle have a primary effect on the magnitude of the eddy current losses.

Dimensional analysis is based on the concept of physically similar systems.[4] If the groupings of parameter values presented in this paper are true correlations and not mere coincidences, then within limits where a correlation holds, fixing the value of either parameter fixes the geometrical pattern of flux change in a lamination at each instant of each cycle. Thus, it can be seen that changing any of the independent variables changes the amount of skin effect. This results in softening the influence on core losses of every one of the classical variables, from what it would be if the flux configuration remained constant.

This softening influence of skin effect can be made apparent by equating the straight-line portion of the correlation in Fig. 1.

$$P_c = \frac{(4.35)B_0 d f^{3/2} J_0^{1/2}}{10^8 \rho^{1/2}} \qquad (1)$$

This equation holds reasonably well over a range of conditions where eddy losses change from about 50 per cent to more than 95 per cent of the core losses.

At the upper end of the range, eddies constitute nearly the entire core loss, and thus approach the condition of varying with 1. the three halves power of frequency; 2. the square roots of hysteresis loss and conductivity; and 3. directly with gauge. The effect of hysteresis loss on eddy currents is a result of its close association with permeability, and is a manifestation of its influence on flux configuration.

At the lower end of the curve of Fig. 1, the eddy currents become independent of the hysteresis loss, because here the influence of skin effect is negligible. Without the softening influence of changing flux configurations, the eddy losses approach the condition of varying with 1. the squares of frequency, flux density, and gauge, and 2. the reciprocal of resistivity.

It is worth noting that over the range of conditions where equation 1 holds, the frequency variation of losses is independent of the hysteresis loss; and this therefore cannot be evaluated by standard core loss separation tests under conditions where equation 1 applies.

All the statements concerning the influence of variables within the range of equation 1 also hold qualitatively for any other range of conditions where core losses are governed by a straight-line parametric relation. The influence of skin effect is more pronounced among materials of group II. For example, the lower portion of the correlation for the lighter gauge materials of group II plotted in Fig. 3 is

$$P_e = \frac{(8.4) f^{7/6} J_0^{5/6} B_0^{1/3} d^{1/3}}{(1{,}000) \rho^{1/6}} \qquad (2)$$

Whether the empirical curves are straight or not, their slopes are such that within limits where the correlations hold, the effects on core losses of the variables other than hysteresis are never more severe than as indicated in equation 1.

Losses Accompanying the Limiting Flux Distribution

If any one of the materials in Figs. 1, 2 and 3 is given a series of tests at successively higher frequencies, eventually the losses begin to increase with the square of the frequency. When this happens, the parameters no longer fit the empirical curve to which the particular material belongs. This failure of the correlation results from the fact that the skin effect increases with increasing frequency until saturation of material near the surfaces of the laminations prohibits the flux configurations demanded by the parametric relation.

It has already been pointed out that, within limits where a correlation holds, fixing the value of either parameter fixes the flux configurations that occur during each cycle, and that this condition is necessary to and responsible for the correlation. It has also been noted that skin effect increases from negligible amounts at small values of the parameters to ever-increasing amounts at higher values. More skin effect means that for a given average flux density, the actual density at the surface of a lamination is higher. It follows that if the combination of skin effect and operating flux density is made sufficiently severe, saturation of part of the lamination will occur, and then the correlation between the parameters will fail.

The authors have developed theoretical eddy current equations based on a limiting flux distribution, and have found that they adequately describe losses when the empirical relations fail because of saturation. This limiting flux distribution represents the maximum possible skin effect, and was first described by A. G. Ganz[1] as a basis for calculating eddy currents in pulse transformers. It consists of regions of saturated material extending inward from the surfaces of the laminations, and separated from material saturated in the opposite direction by surfaces parallel to the lamination surfaces. Magnetization proceeds inward by movement of the separating walls. Simple mathematical analysis is possible to determine eddy currents accompanying such changes in magnetization. The derived equation for average alternating current eddy losses accompanying this limiting flux distribution is as follows

$$P_e = \frac{(\pi)^2}{(4)(10^{16})} \frac{f^2 B_0^2 d^2}{\rho} \frac{B_0}{B_a} \qquad (3)$$

where

P_e = eddy current losses (watts per cubic centimeter)

B_a = available flux (95 per cent of static induction at 10 oersteds)

The definition of the available flux here is necessarily somewhat arbitrary.

Brailsford[5] has pointed out that skin effect may give increased hysteresis loss. The hysteresis losses accompanying this limiting flux distribution would be

$$P_h = f J_a \frac{B_0}{B_a} \qquad (4)$$

where J_a is the static hysteresis loss

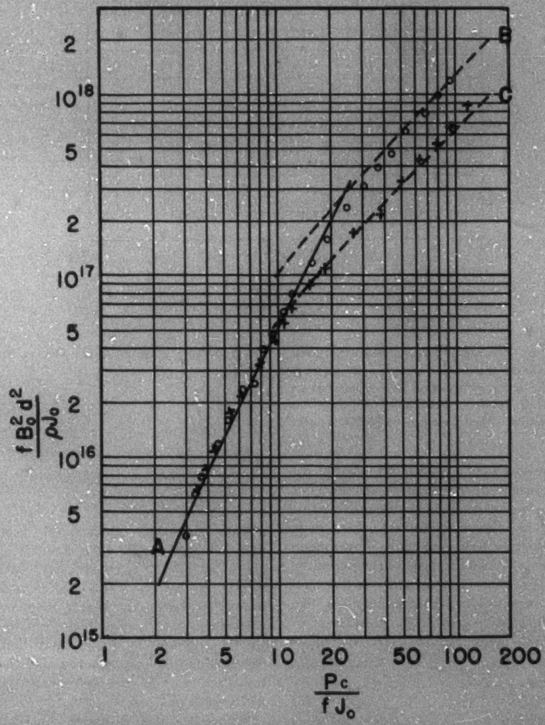


Fig. 5. Theoretical limiting losses in 15-mil oriented silicon steel. Curve A is the empirical relation from Fig. 1. Curves B and C are values of equation 7 for oriented silicon steel at 5,000 and 10,000 gausses respectively. Circles are measured values at 5,000 gausses, and crosses are measured values of the parameters at 10,000 gausses

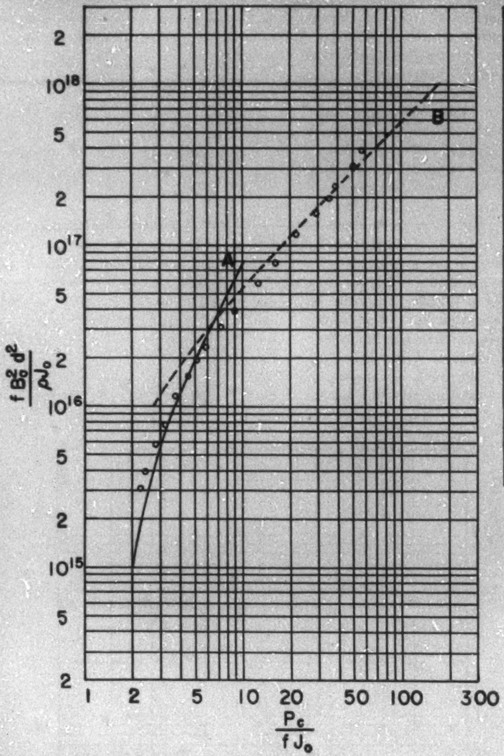


Fig. 6. Theoretical limiting losses in 10-mil Deltamax. Curve A is from Fig. 2. Curve B is the locus of equation 8 for Deltamax at 10,000 gausses. The circles are measured values for 10-mil Deltamax at 10,000 gausses

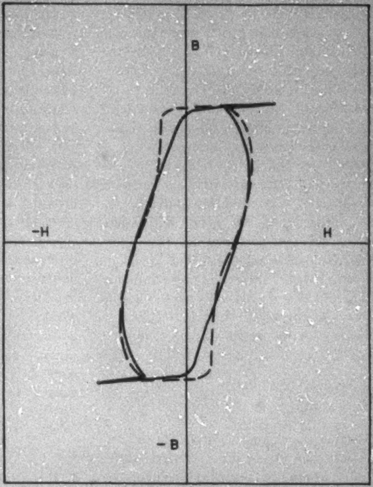


Fig. 7. A comparison of the shapes of an experimental a-c hysteresis loop (broken line) and a theoretical one (solid line). These are for 10-mil Deltamax operating at 400 cycles per second

measured from a peak flux density of B_a. Because hysteresis is only a small part of the core losses under conditions where this limiting flux distribution exists, it is sufficiently accurate to consider the static hysteresis losses in group II materials as varying with the first power of flux density, and in group I materials as varying with the 1.6 power of flux density.

This gives for hysteresis losses in group I

$$P_h = fJ_0\left(\frac{B_a}{B_0}\right)^{0.6} \tag{5}$$

and in group II

$$P_h = fJ_0 \tag{6}$$

Then, taking the total losses as the sum of hysteresis and eddies, we have for group I materials

$$\frac{P_c}{fJ_0} = \left(\frac{B_a}{B_0}\right)^{0.6} + \frac{(\pi)^2}{(4)(10^{16})}\,\frac{fB_0^2d^2}{\rho J_0}\,\frac{B_0}{B_a} \tag{7}$$

and for group II materials we have

$$\frac{P_c}{fJ_0} = 1 + \frac{(\pi)^2}{(4)(10^{16})}\,\frac{fB_0^2d}{\rho J_0}\,\frac{B_0}{B_a} \tag{8}$$

Comparison of experimental losses with values from equations 7 and 8 are given for 15-mil oriented silicon steel in Fig. 5, and for 10-mil Deltamax in Fig. 6. It can be seen that as frequency is increased in each case, the losses follow the appropriate empirical curve until the limiting loss is approached, and that they then begin to increase in accordance with equations 7 and 8.

Materials operated under conditions where this limiting flux distribution prevails show distinctively shaped dynamic hysteresis loops as seen in Fig. 7. This oscillogram was supplied by H. W. Lord, who first noted and called the authors' attention to these loops. It was taken on 10-mil Deltamax at 400 cycles per second. The same analysis used in deriving equation 3 gives the following equation for the instantaneous magnetizing force

$$H_e = \frac{(\pi^2)fd^2B_0^2(\cos\omega t)(1-\sin\omega t)}{(10^9)\rho B_a} \tag{9}$$

where

H_e = magnetizing force of eddies, oersteds
$\omega = 2\pi f$
t = time, seconds

This equation holds only over the range of $\omega t = 90$ degrees to $\omega t = 270$ degrees, and where the instantaneous flux density is $B = B_0 \sin \omega t$.

A comparison of the shapes of the experimental loop and a synthetic one made as the sum of the static loop and calculated eddy currents is given in Fig. 7. The corners of this loop indicate that the limiting flux configuration is not reached immediately at the start of each half-cycle, and this results in losses slightly higher than theoretical.

Moly-Permalloy and Mumetal both show measured limiting losses which are about 20 to 30 per cent higher than the losses calculated from equation 7. This is partly because the residual induction (B_r) in each of these metals is only 70 per cent of the available induction (B_a).

For these materials, substitution of

$$B_a' = \frac{B_r + B_a}{2}$$

for B_a in equation 7 gives closer agreement with measured values.

The effects of material variables on eddy losses accompanying this limiting flux configuration are evident in equation 3. Because frequency gauge, and resistivity do not influence the flux distribution, their effects are neither softened nor amplified. However, since the relative depth of flux change is proportional to the operating flux density and inversely proportional to the available flux, the losses vary with the cube of the operating density instead of the square, and vary inversely with the available flux. This latter gives another possible control over eddy losses through better materials.

Domain size is not a factor determining losses accompanying this limiting flux configuration, because magnetization proceeds along moving walls which are equivalent in every material, regardless of the basic domain structure.

The existence of this limiting flux distribution demonstrates that when the pattern of flux change in a magnetic lamination is known, a-c core losses can be calculated with good accuracy. This is in agreement with the findings of Williams, Shockley, and Kittel[2]. Thus, it tends to corroborate their idea that classical eddy current equations fail because they ignore domain configurations; and it tends to strengthen the idea offered in this paper that the unknown material variable found in the empirical plots is domain size.

Conclusions

When the combined effect of frequency, gauge, flux density, permeability, and conductivity is sufficiently severe, the configuration of flux change in a lamination reaches a limiting condition representing the maximum skin effect. The losses which accompany this limiting flux distribution agree well with calculated values.

When flux distribution is not limited by saturation, the core loss behaviors of the materials tested fall into two groups. Members of each group are characterized by similarly shaped magnetization curves.

Within each group, significant correlations have been found between parameters embodying the classical factors known to influence core losses. However, there exists in each group an additional material variable; and it is suggested that this is domain size.

These correlations suggest that significant regularity of effective domain size exists within groups of magnetic materials.

Test data indicate that core losses in modern magnetic materials would be decreased 40 per cent or more by reducing the effective domain size to that which exists in ingot iron, while keeping other factors the same.

References

1. APPLICATIONS OF THIN PERMALLOY TAPE IN WIDE-BAND TELEPHONE AND PULSE TRANSFORMERS, A. G. Ganz. *AIEE Transactions (Electrical Engineering)*, vol. 65, April 1946, pp. 177–83.

2. STUDIES OF THE PROPAGATION VELOCITY OF A FERROMAGNETIC DOMAIN BOUNDARY, H. J. Williams, W. Shockley, C. Kittel. *Physical Review*, New York, N. Y., vol. 80, no. 6, Dec. 15, 1950, pp. 1090–94.

3. LOSSES IN ELECTRICAL SHEET STEEL, K. H. Stewart. *Journal*, Institution of Electrical Engineers, London, England, vol. 97, pt. II, 1950, pp. 121–25.

4. ON PHYSICALLY SIMILAR SYSTEMS; ILLUSTRATIONS OF THE USE OF DIMENSIONAL EQUATIONS, E. Buckingham. *Physical Review*, New York, N. Y., vol. 4, Oct., 1914, pp. 345–77.

5. INVESTIGATION OF THE EDDY CURRENT ANOMALY IN ELECTRICAL SHEET STEELS, F. Brailsford *Journal*, Institution of Electrical Engineers, vol. 95, pt. II, 1948, pp. 38–48.

◆

No Discussion

Electron-Tube Voltmeter Concepts in the Mid-Frequency Range

GEORGE B. HOADLEY
MEMBER AIEE

FROM TIME to time the AIEE sponsors papers of a tutorial nature designed to expound some phase of electrical engineering for the benefit of the general reader rather than for the specialist. This paper is of the tutorial type, treating some of the problems in electron tube voltmeter design and operation, without trying to be all inclusive. The term mid-frequency range is used to indicate that none of the effects which set the upper and lower frequency limits of the instruments are considered.

General Discussion

The electron tube presents a special challenge to those interested in instruments for making measurements of electrical quantities because the electron tube is not, in itself, a precision device. Suitable circuits must be designed to allow the quick response and low input-current characteristics of the tube to be used to advantage without incurring serious limitations on accuracy and permanency of the calibrations.

The most important means of stabilizing electron-tube circuit performance is that of negative feedback. This basic principle has received much attention in recent years not only as applied to all-electric circuits, but also to electromechanical circuits, where it is used in the study of servomechanisms.

Feedback Amplifier Voltmeter

A most useful feedback amplifier circuit is shown in Fig. 1. Here the first stage of amplification is shown, with a cathode resistor R. This resistor carries not only the cathode current of the first tube, but also the output current of the amplifier. This output current I_{out} flows through this resistor on its way to the common cathode-return conductor of the amplifier. Some of this current may also go through the first tube, but for the sake of simplicity we will assume that R is low enough so that this effect is negligible. Also, we will assume that the signal current of the first tube is negligible compared to the output current I_{out}.

In passing through this resistor R, the output current causes a rise in voltage from C to K, which voltage must be combined with the input voltage V_{in} to give the voltage V_{KG} which actuates the amplifier and determines the value of output current I_{out}.

To analyze this situation, three simultaneous equations are written, expressing

---

Paper 53-308, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 17, 1953; made available for printing May 26, 1953.

GEORGE B. HOADLEY is with the North Carolina State College, Raleigh, N. C.

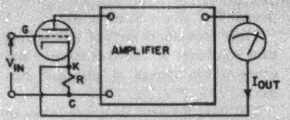


Fig. 1. A common circuit for amplifier stabilization by feedback

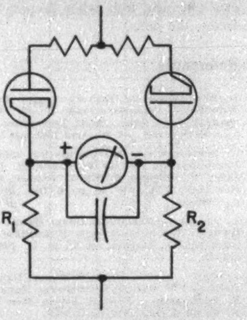


Fig. 2. A full-wave rectifier for an electron-tube voltmeter circuit

in mathematical terms the ideas which have just been described. These equations are

$$V_{CK}=I_{\text{out}}R \tag{1}$$

$$V_{\text{in}}=V_{CK}+V_{KG} \tag{2}$$

$$I_{\text{out}}=gV_{KG} \tag{3}$$

If the second equation is rewritten by factoring V_{CK} out of the right-hand side, there results

$$V_{\text{in}}=V_{CK}\left(1+\frac{V_{KG}}{V_{CK}}\right) \tag{4}$$

Now equations 1 and 3 are solved to find the voltage ratio

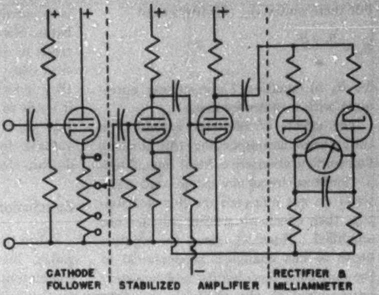
$$\frac{V_{KG}}{V_{CK}}=\frac{1}{gR} \tag{5}$$

Substituting equations 1 and 5 into equation (4) there results

$$V_{\text{in}}=I_{\text{out}}R\left(1+\frac{1}{gR}\right) \tag{6}$$

If gR is made considerably larger than 1, the output current I_{out} multiplied by the resistance R is the input voltage with an error expressed as $100/gR$ per cent. The most important thing about this result is that the characteristic g which is the overall transconductance of the amplifier and is the only term involving the characteristics of the electron tubes, appears only in the error term! All that is required of the amplifier itself is that it amplify enough. Then g will be large enough so that $1/(gR)$ will be negligible.

Fig. 3 (right). An electron-tube voltmeter circuit



In a practical case, R may be 10 ohms. With a 3-stage amplifier comprising two voltage stages each having a gain of 50, and an output stage having a transconductance of 5000 micromhos, the overall transconductance g is $50\times50\times5{,}000\times10^{-6}$ which is 12.5 mhos. The term $1/(gR)$ is then $100/(12.5\times10)=0.8$ per cent. This means that the circuit of Fig. 1 can be made into a voltmeter by using an a-c instrument to measure the output current I_{out} but calibrating it in volts input according to equation 6.

Now, if each tube of the amplifier suffers a 10-per-cent loss in its amplifying abilities, g becomes $45\times45\times4{,}500\times10^{-6}=9.1$ mhos and the term $100/(gR)$ is 1.10 per cent. Therefore, if in the original calibration, a nominal value, such as 0.01 had been used for $1/(gR)$, the resulting overall instrument would be within a tolerance of ±0.2 per cent for a 10-per-cent variation in each stage of the amplifier. This is more than sufficient accuracy for many measurements, so we have a satisfactory electron tube voltmeter, having the high input impedance characteristics of the electron tube, coupled with a freedom from sensitivity to the normal variations in electron-tube characteristics.

In the design of the amplifier itself, there are many problems typical of electronic circuits which must be solved. In particular, the amplifier must be so designed at both its high- and its low-frequency limits that the system will not oscillate. This often requires special attention to the frequency characteristics of the amplifier in these regions, even though they are outside the range of operation of the voltmeter.

The indicator used in the output circuit of Fig. 1 may be any type of alternating-current instrument. For power frequencies, an iron-vane or dynamometer movement could be used, although this is seldom done. For audio work, a dry-rectifier instrument would be applicable, and for higher frequencies a thermocouple instrument could be used. But none of these are as generally satisfactory as the combination of a vacuum-diode rectifier and a d-c milliammeter, because none can match its combination of ruggedness, quickness of response, and large frequency range.

One arrangement for such a combination is shown in Fig. 2, which shows a circuit designed to carry the a-c component of output current. Here a bridge rectifier circuit is used so that current will flow during each half of the cycle. The left-hand diode conducts during one half of the cycle. The current divides between the resistors R_1 and R_2, with the current in R_2 passing through the d-c milliammeter from left to right. During the other half of the cycle, current passes upward through the right-hand diode, again with a division of current between R_1 and R_2, but with the milliammeter again receiving current from left to right. Consequently, the milliammeter receives a unidirectional current, with its variations smoothed out by the shunting capacitor. Thus its reading is propor-

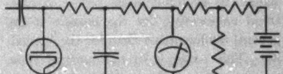


Fig. 4. A half-wave rectifier with zero suppression for a logarithmically scaled instrument

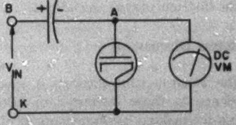


Fig. 5. The basic peak-reading diode circuit

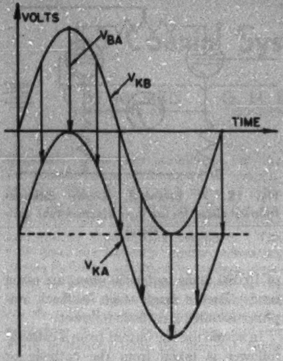


Fig. 6. The suppression of the voltage across the diode of Fig. 5

tional to the average of the rectified input voltage to the overall voltmeter. Usually the milliammeter is calibrated in terms of rms input volts for a sinusoidal input. Then if the input wave contains harmonics the voltmeter reading will not give the rms value, but will give an indication nearly proportional to the average of the rectified wave.

In some circuits, a cathode follower, which will be described later, is placed at the input of the voltmeter. When used with the feed-back amplifier type of voltmeter, a change of range can be accomplished by connecting the input of the feedback amplifier to a tap on the cathode resistor of the cathode follower. This, allows a wide range in input voltages, since the cathode follower can accomodate a large input. A simplified schematic diagram of a voltmeter of this type is shown in Fig. 3.

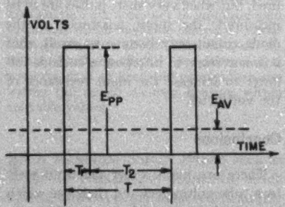


Fig. 7. Voltage pulses used in the analysis of the peak-reading diode operation

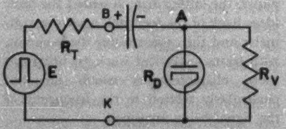


Fig. 8. The equivalent circuit of the peak-reading diode connected to a circuit for measurement of voltage

Another circuit, shown in Fig. 4, employs half-wave rectification and a circuit for injecting a small negative direct current into the milliammeter so that the lowest reading on the scale will be 1/10 of the highest reading instead of zero. This aids in obtaining a logarithmic scale. When this circuit is used in place of the circuit of Fig. 2 as the output indicator in the circuit of Fig. 1, the feedback causes an improvement in the linearity of the rectifier.

Peak Diode Voltmeter

A totally different type of electron tube voltmeter is built around the peak-reading diode circuit, which is shown in its elements in Fig. 5. When this is connected to a circuit for measurement, current flows in the diode whenever the anode is more positive than the cathode. At first, this occurs for each half cycle, but the diode current charges the capacitor, so that diode current can flow only when the input voltage V_{KB} is more positive than the voltage V_{AB} across the capacitor. Consequently, diode current flows during less and less of the cycle of the voltage being measured until in the perfect case, V_{AB} has become equal to the positive peak value of V_{KB}, which is the input voltage. Then the three voltages appear as they are shown in Fig. 6, with the voltage across the diode being the input voltage depressed until it is completely negative, just touching the zero line when the input voltage has positive peaks.

The d-c voltmeter in shunt with the diode will read the average of the voltage across the diode, which will be the difference between the average value and the positive-peak value of the input voltage, and so this ideal circuit reads the positive-peak value of the input.

This analysis neglects the current in the d-c voltmeter, and if the errors are to be studied, this current must be considered. The direct current through the d-c voltmeter must be supplied by the diode, so the diode must conduct during part of each cycle. Consequently, the voltage V_{KA} must become positive at some part of the cycle so the depression of the input voltage is not complete, and some error will result.

In order to analyze this, it is easiest to discard the sinewave input and to use a square topped, pulsed input, such as is shown in Fig. 7. From this figure, the average value of E is readily calculated to be

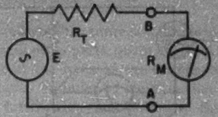


Fig. 9. A voltmeter connected to a circuit with internal resistance

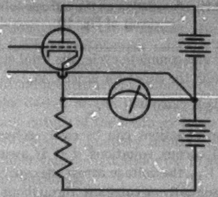


Fig. 10. A simple cathode-follower circuit for use as the d-c voltmeter of Fig. 5

$$E_{AV}=E_{pp}\frac{T_1}{T} \tag{7}$$

and the reading of the voltmeter according to the simplified theory would be

$$\text{Theoretical reading}=E_{pp}-E_{AV}=E_{pp}\frac{T_2}{T} \tag{8}$$

The actual reading, however, will be the average current through the voltmeter resistance R_v, multiplied by R_v. Inasmuch as the average current through the capacitor is zero, this will be the same as the average current through the diode, multiplied by R_v.

Analysis of this action will be carried out with reference to Fig. 8, which shows a diode circuit connected to an electromotive force E in series with an equivalent Thevinin resistance R_T. This is done so that the result of the analysis will yield information relative to the loading effect of the voltmeter on the circuit being measured.

By a study of this circuit and Fig. 7, we can express the current in the capacitor for each part of the cycle. The current through the capacitor when the input voltage is positive is

$$I_+=\frac{E_{pp}-V_{AB}}{R_T+\dfrac{R_vR_d}{R_v+R_d}} \tag{9}$$

When the input voltage is zero, the capacitor current is in the reverse direction and is

$$I_-=\frac{-V_{AB}}{R_T+R_V} \tag{10}$$

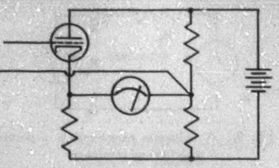


Fig. 11. A variation on the cathode-follower circuit of Fig. 10

Over one complete cycle, the total charge through the capacitor is zero, so

$$I_+T_1+I_-T_2=0 \qquad (11)$$

These equations can be easily solved as simultaneous equations if it is assumed that V_{AB}, the voltage across the capacitor, is essentially a constant throughout the cycle. This means that this development is good only if C is large enough so that the time constants are large compared with T_1 and T_2.

By eliminating I_- and V_{AB} from these three equations, there results

$$I_+\left[R_T+\frac{R_vR_d}{R_v+R_d}+\frac{T_1}{T_2}(R_T+R_v)\right]=E_{pp} \qquad (12)$$

The diode current is I_+ multiplied by the proper resistance ratio, and the average of this is

$$I_{D_{av}}=I_+\frac{R_v}{R_d+R_v}\,\frac{T_1}{T} \qquad (13)$$

so that the actual reading of the instrument is

$$\text{Reading}=I_{D_{av}}R_v=I_+\frac{R_v^2}{R_d+R_v}\,\frac{T_1}{T} \qquad (14)$$

When this is combined with equations 12 and 8, there results, after some algebraic manipulation

$$\text{Theoretical Reading}=\text{Actual Reading}\times\left[1+R_T\frac{T}{T_1}\left(\frac{R_d+R_v}{R_v^2}\right)+\frac{R_d}{R_v}\,\frac{T}{T_1}\right] \qquad (15)$$

which is the result sought.

The last term in the brackets gives the error resulting from the finite resistance of the d-c voltmeter. If the pulse width, T_1 is only 1 per cent of the period T, then R_v must be 10,000 times R_d for this error to be 1 per cent. If the diode resistance when conducting is 500 ohms, then R_v must be 5 megohms or more, which means that something special must be done with the d-c voltmeter.

The second term in the brackets can best be interpreted in terms of an equivalent resistance of the voltmeter. To do this, consider the simple circuit of Fig. 9. Here a voltmeter is measuring the voltage between terminals A and B, and from its reading there can be calculated the voltage between A and B before the voltmeter was connected. This gives

$$E=\frac{\text{Actual Reading}}{R_M}(R_M+R_T) \qquad (16)$$

or

$$E=\text{Actual Reading}\times\left[1+\frac{R_T}{R_M}\right] \qquad (17)$$

By comparison with equation 15, the effective input resistance of the diode circuit is the reciprocal of all of the second term in the brackets, except for R_T. This gives

$$R_M=\frac{T_1}{T}\,\frac{R_v^2}{R_d+R_v}=\frac{T_1}{T}\,\frac{R_v}{1+\dfrac{R_d}{R_v}} \qquad (18)$$

Using the values as before, for an example, the input resistance is

$$R_M=0.01\,\frac{5\text{ megohms}}{1.0001}=50{,}000\text{ ohms} \qquad (19)$$

This is, of course, an equivalent resistance. During part of the cycle, the resistance is less than this, and during another part of the cycle, it is greater than this, but this value can be used with the simple circuit of Fig. 9 to calculate the effect of the voltmeter on the voltage being measured.

Both of the errors shown in equation 15 depend on the time ratios of the applied electromotive force. Consequently, it would be expected that the equations for a sinusoidal applied voltage would differ somewhat from these. The same basic ideas would be used for the development of the theory, but the mathematics would be more involved, partly because the times at which conduction in the diode started and stopped would not be known, but would have to be found.

The Cathode Follower

The analysis just finished has shown the R_v must be of the order of 5 megohms for satisfactory freedom from errors in the diode circuit. Usually this means that some form of electron tube amplifier must be used to measure the d-c component of the voltage across R_v. Here again the feedback amplifier is of use, but usually only one stage of amplification is employed, with the plate current of the first tube acting as the I_{out} in Fig. 1. Then the resistance R must be much higher than was used before, because g is much lower, so values of R of the order of 10,000 ohms to 50,000 ohms are often used. Such a single-stage feedback amplifier is called a cathode follower.

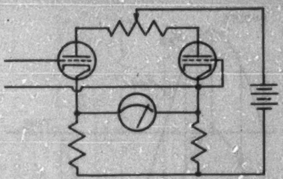


Fig. 12. A balanced, 2-tube cathode-follower circuit for use in an electron-tube voltmeter

In actual use the output from a cathode follower is taken from the cathode or from a tap on the cathode resistor R with a return to the common terminal. Sometimes this return is through an auxiliary circuit designed to eliminate the no-signal d-c component. Thus, in Fig. 10, the return is to a tap on the plate supply of the tube. In Fig. 11, there is shown an alternative in which the return is connected to a tap on a resistive voltage divider placed across the supply. Still another alternative is to use a second cathode-follower circuit, as shown, in Fig. 12. In all of these, various adjustments are provided to allow complete elimination of the no-signal component. The circuit of Fig. 12 relies on similarity between the two halves of the circuit to provide stability of the circuit against variations in ambient conditions.

Usually, the peak-reading diode is connected directly to the circuit being measured, but when very short pulses are to be measured, the input resistance of the diode circuit may become so small, that it is necessary to interpose a cathode follower to increase the input resistance of the voltmeter.

Conclusions

There are many other facets of electron tube voltmeters and their use which have not been touched upon in this paper. Such matters as frequency response, logarithmic scales, and drift compensation could receive attention. In this paper, the author has presented the basic ideas back of the feedback-amplifier voltmeter and the peak-reading diode circuit, with some analytical work to bring out those characteristics which are not particularly related to the frequency of the signal being measured.

No Discussion

L3 Coaxial System—Amplifiers

L. H. MORRIS
NONMEMBER AIEE

G. H. LOVELL
NONMEMBER AIEE

F. R. DICKINSON
NONMEMBER AIEE

Synopsis: The line amplifiers for the *L3* coaxial system are designed to compensate for the loss of the 4 miles of cable which separate the repeaters; the flat amplifiers are used to compensate for equalizer loss and as transmitting amplifiers. The two types are basically similar, consisting of two feedback amplifiers in tandem, separated by an interamplifier network; in the line amplifier, this network is variable, and is automatically adjusted to compensate for variations in cable temperature, and for small deviations from the nominal 4-mile spacing.

Coupling networks, employing high-precision transformers, are used to connect the amplifiers to the coaxial cable through the required power separation filters. The low-impedance windings of the transformers are center-tapped and a balancing network provided to match the cable impedance over the transmitted frequency band. The amplifiers are equipped with plug-in tubes of high figure-of-merit, which were developed for this application. A double-triode output stage is used to obtain improved system signal-to-noise performance. Provision is made for preventive maintenance of vacuum tubes and for a controlled adjustment of gain on an in-service basis.

All important components of the amplifier are subject to quality-control procedures to assure that the average gain of groups of amplifiers will be held within narrow limits and that individual amplifiers will form a normal distribution around the average. This approach is essential in order to meet system equalization and signal-to-noise objectives. Careful mechanical design and rigid control of the mechanical aspects of manufacture are necessary to minimize gain variations which might be caused by variations of parasitic circuit elements and unwanted feedback effects. Special measures were required to keep the temperature rise within the sealed die-cast housing within tolerable values.

IN A TRANSMISSION system such as the *L3* coaxial, the degree to which system objectives are achieved is largely dependent on the quality of the amplifiers which compensate for the cable loss. To a considerable extent, the same statement applies to the similar flat-gain amplifiers used to make up for the loss of the equalizers at various points along the route. The development of an amplifier which would meet the exacting requirements of the *L3* system was, in turn, dependent on new developments in the fields of vacuum tube design and circuitry network-design techniques, element and network fabrication, and statistical quality control. To these new tools were added the lessons learned in years of manufacture and operation of the preceding *L1* system.

The importance of some of these factors can best be illustrated by examining the implications of the amplifier requirements which follow from the material in the companion papers on system design[1] and equalization.[2] Obviously the figure-of-merit, modulation coefficients, and life of the vacuum tubes will be determining factors in setting the amount of feedback that can be obtained, and the signal-to-noise ratio of the system. It is not so immediately apparent that system requirements could not be met with the present 4-mile repeater spacing, if it were not for the use of quality control at every stage of manufacture from elements, and even the raw materials entering into components, to completed amplifiers. As the companion papers[1,2] show, however, the equalization plan of the system is predicted on a degree of reproducibility of amplifier gain and delay characteristics obtainable only by quality control applied at every stage of the manufacturing process.

The present equalization plan is based on the assumptions that the gain of the average line amplifier will match the loss of the preceding line section to within 0.15 decibel (db), and that the average amplifier gain will not vary from one batch of new amplifiers to another by more than about 0.06 db. Under these assumptions, a system equalization plan can be worked out which results in reasonable spacing between equalizers, a tolerable signal-to-noise penalty due to misalignment and equalizer loss, and a practicable procedure for adjusting long systems. Any gross departure from the basic assumptions as to reproducibility of amplifiers would seriously compromise these objectives, which even now are achieved only by using equalization which requires a flat-gain amplifier for every four or five line amplifiers. Now it turns out that with the most precise elements that can be made, the gains of individual amplifiers will vary by about ±0.6 db. We need, therefore, a tool which will permit us to control the gain of the average amplifier to an order of magnitude of greater precision than we can economically control the individual. This is exactly the effect which modern methods of statistical quality control aim at achieving, and since the entire amplifier is merely the sum of its parts, quality control must start at the roots of the manufacturing process. In order to apply quality control intelligently, and to be sure that all important causes of gain variation are understood, it has been necessary to carry out, side by side with empirical laboratory work, a program of computing the insertion gain of the amplifier, starting with fundamental element values. These computations have also proved of value in obtaining satisfactory stability margins in the design of the low- and high-frequency cutoffs of the feedback loops, and in obtaining preliminary information on amplifier-gain deviations for use in equalization planning.

The severe gain and delay reproducibility objectives also have their effect on the mechanical design of the amplifier. At first sight, the unit appears to be a lumped-constant structure rather than one in which distributed effects would be of paramount importance. Usually the circuit designer in such a case is interested in the mechanical design only for reasons of neatness and economy, but when we look deeper we find that in the amplifier structure as a whole, as well as in the case of certain components, the effects of distributed capacitance and inductance could easily defeat our objectives if the mechanical design were not such as to assure precise control of element placement and wiring lengths.

For these reasons, an order of mechanical accuracy is specified beyond that which can be justified by the accuracy of transmission measurements on individual amplifiers. This striving for mechanical accuracy is carried all the way through, from element piece parts, through subassemblies to the assembly on the final amplifier framework. The logical expectation is that by reproducing the amplifiers as exactly as possible, we will control not only the known elements but also the many parasitic effects, and thereby minimize the appearance of

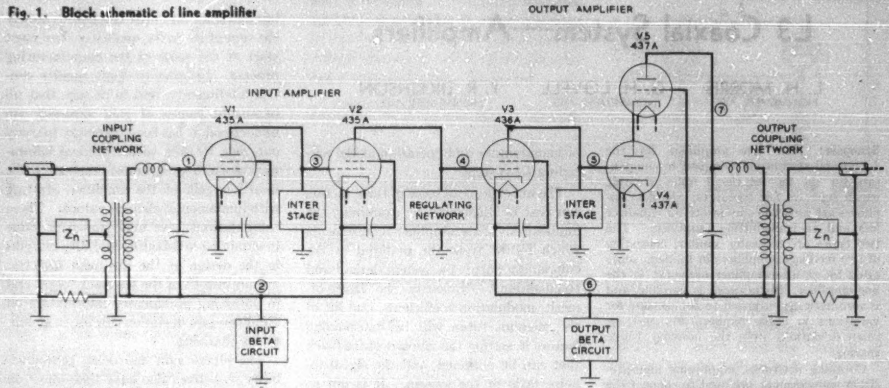
---

Paper 53-222, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 18, 1953; made available for printing April 27, 1953.

L. H. Morris, G. H. Lovell, and F. R. Dickinson are with the Bell Telephone Laboratories, Inc., Murray Hill, N. J.

Many members of the Bell Laboratories have made important contributions to the *L3* amplifier design. Particular mention should be made of the work of S. E. Miller on fundamental amplifier design, S. Darlington and T. R. Finch on network design, C. W. Thulin on the precision transformer, B. J. Kinsburg on quality control problems, E. Ley on mechanical design, and E. F. O'Neill on the flat-gain amplifier and on high-frequency stability problems in both amplifiers.

Fig. 1. Block schematic of line amplifier



shifts in amplifier gain and delay, which would be substantial in the system, even though difficult to detect in the measurement of individual amplifiers.

Another design feature of the amplifier, based on the system equalization point of view, is the omission of all adjustable elements, or trimmers, to control the transmission. By eliminating gain adjustments, the possibility of adjusting one element to compensate for the shortcomings of another, and the possibility of systematic errors of setting due to faulty or inaccurate adjustment techniques, are both eliminated. Either of these possibilities would tend to convert relatively large random effects into smaller but systematic effects, a conversion which would penalize the system equalization. Obviously, however, the elimination of gain-adjusting elements puts a higher premium than ever on the requirement that all elements, including tube capacitances, show only small random variations about tightly controlled nominal values.

In addition to the gain requirements discussed above, the following line-amplifier objectives are to be met:

1. The gain of the amplifier must be continuously variable, under the control of a regulator circuit, to compensate for differences in lengths of cable sections, and for variations of cable loss caused by temperature changes. The shape of the gain change should match the square root of frequency shape of the line-loss change over the transmitted band to within a few hundredths of a decibel. This accuracy of shape, which, of course, is based on equalization considerations, should hold over the entire regulation range, which is about ±6 db at the top transmitted frequency.

2. The input and output impedances of the amplifier should match the cable impedance in order to minimize the effects on television transmission of echoes caused by line irregularities, such as splices. Tolerable values of reflection coefficient at amplifier input and output are about 5 per cent at the television carrier and 10 to 15 per cent at the upper-band edge.

3. Feedback consistent with system-modulation requirements, shaped across the transmitted band to minimize low-frequency intermodulation products and to give a smooth, easily equalized shape of gain change as tubes age, must be obtained, while maintaining adequate margins against singing.

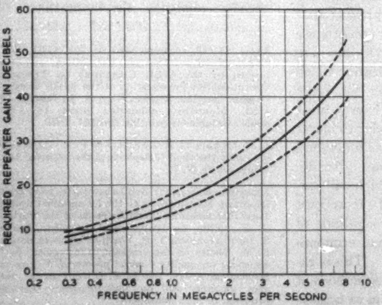


Fig. 2. Required line repeater gain

4. Other requirements include design and selection of elements to assure as small a change of gain versus ambient temperature as possible, mechanical design provisions for keeping the temperature rise within the unit to a minimum, both in order to keep the temperature-gain effect small and to obtain long element life, a sealed housing to avoid damage by humidity in exposed locations, and provision of facilities for testing tubes in service.

Configuration

The circuit configuration of the line amplifier is shown in Fig. 1. It consists of two independent feedback amplifiers, with a regulating network between them acting like a 2-terminal interstage. Each of the amplifiers is essentially a 2-stage circuit—the input amplifier, literally so and the output amplifier essentially so, since the double-triode output circuit acts like a single stage. Each amplifier is connected to the coaxial through a coupling network, which consists of a transformer plus gain-shaping and impedance-adjusting elements. Since there is no feedback around the coupling networks, they directly affect the insertion gain of the amplifier, as do the two beta circuits and the regulating network. The required shaping of insertion gain across the transmitted band is obtained and controlled by the design of these five networks. Fig. 2 shows the required amplifier gain for nominal and extreme thermistor settings; these required gains differ from the line loss by the small losses of the associated repeater components. At a midrange thermistor setting, about 37 db of gain shaping is needed. The manner in which this shaping is distributed among the five networks has important effects on the feedback which can be obtained, and the sensi-

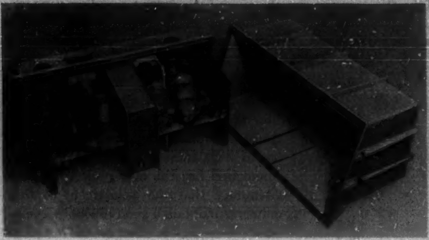


Fig. 3. Line amplifier and sealed housing

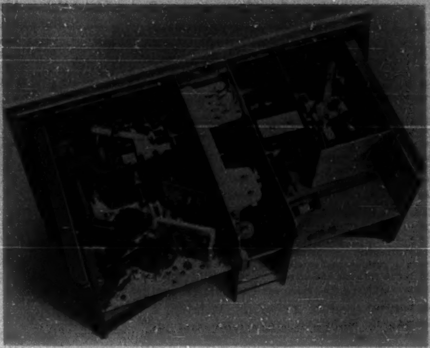


Fig. 4 (right). Another view of amplifier, showing networks

tivity of amplifier gain to element variations.

This configuration offers several advantages, one of the most important of which is that the regulating network is between the amplifiers. In most other configurations, the only gain-determining network available for the regulating function is the beta circuit. When the feedback is not infinite, this introduces errors for which it is difficult to compensate, and limits the available feedback by complicating the design. In this configuration, the impedances which the amplifiers effectively present as shunts across the regulating network are high, and can be allowed for in the design, so that the regulation error can be made much smaller than the error associated with beta-circuit regulation in an amplifier having relatively little feedback. Other major advantages are the superior signal-to-noise performance, and the relative simplicity of the feedback loops of this amplifier, as compared to alternative designs.

Mechanical Design

The mechanical assembly of the amplifier, like the configuration, is divided into three main sections. The input and output amplifiers are mounted on separate chassis, which are designed for ready removal from the main base casting. The regulating network is mounted in an enclosed, shielded compartment, which also serves to shield the component amplifiers from each other. Figs. 3 and 4 show the assembled amplifier.

Because of the wide frequency range and close control of parasitic capacity and lead inductance required, the *L3* amplifier was designed as an integrated whole, and all networks were designed with, and as part of, the amplifier. In each case circuit elements were placed in space in the best possible position for optimum electrical performance, and supporting structures were then designed to maintain the desired space relationships. These supporting structures are made as separate units which mount on the amplifier chassis, so that the networks can be individually tested before final assembly into the amplifiers, and can be removed for repair or replacement if necessary.

Heretofore, this method of design has been impracticable because it results in very complicated supporting structures. It was feasible in this case because of the availability of a new type of material, which removes many of the mechanical-design constraints. This is a cold-casting resin, and parts are produced in a cheap phenolic mold. Since the process is practically equivalent to the sand-casting of metals, complex parts can be economically manufactured even in the relatively small quantities required for *L3* production. Where necessary to assure accurate location or orientation of parts, the cast resin frameworks are milled or spot-faced. The structure used in the input interstage, shown in Fig. 5, is an example. Circuit elements are wired to pins driven into the casting or, in some cases, to wires imbedded in the material. All wiring can thus be made direct with a minimum of bending and no doubling back, resulting in a reproducible and uniform product.

The entire ampifier is housed in a sealed die-cast container to protect the components from humidity damage. Dessicant is enclosed in each amplifier. It would have been desirable to make this housing of aluminum, but the high melting point of this metal makes it impractical for such a large casting to meet the air-tightness requirement. A zinc-base die casting alloy was therefore used. Sealing is accomplished by rubber gaskets at all openings for connections, and at the joint between the two parts of the housing. The removable part of the casting which serves as a cover is made as large a part of the total housing as possible, in order to provide maximum accessibility for maintenance of the units mounted on the base. The entire housed unit is arranged to mount on a relay-rack mounted panel by means of slides which are self-locking. Signal and power connections are made by means of flexible cords which are available on the panel to plug into the unit after it is in place on the slides.

Vacuum Tubes

The tubes have been fully described in an earlier paper[3]; their characteristics are summarized in Table I. They are plugged into conventional sockets, and single rather than parallel tubes are used in each stage. These are departures from the practice of the *L1* coaxial system, in which two tubes in parallel were soldered in each stage. The use of sockets increases the parasitic capacities and reduces the obtainable feedback by 1 or 2 db, but it was felt that the resulting maintenance economy was worth this sacrifice. With single tubes, the failure of one tube results in a line failure and a

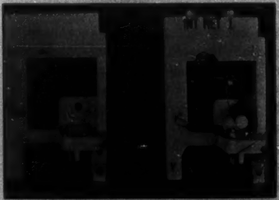


Fig. 5. The input interstage, illustrating use of cast resin frameworks

Table I. Tube Characteristics: Design Center Values, New Tubes

	435A Tetrode	436A Tetrode	437A Triode
Heater, voltage, volts	6.3	6.3	6.3
Heater current, amperes	0.3	0.45	0.45
Plate-cathode voltage, volts	180	180	150
Screen-cathode voltage, volts	150	150	
Grid-ground voltage, volts	9.0	9.0	9.0
Cathode bias resistor, ohms	630	315	262
Plate current, milliamperes	13.1	23.4	40.2
Screen current, milliamperes	3.2	8.6	
Transconductance, milliamperes per volt	16.5	32	47
Plate resistance, ohms			970
Grid-cathode bias, volts	1.3	1.1	1.5
Capacitances (Approximate hot values in micromicrofarads)			
Grid to plate	0.02	0.04	3.6
Grid to cathode and screen	9.2	18.2	16.4
Grid to heater	0.6	0.8	0.1
Plate to cathode and screen	1.0	1.5	0.7
Heater to cathode and screen	5.4	7.2	5.4
Heater to plate	1.7	1.9	0.2
Modulation*			
Second order, decibels	36		36
Third order, decibels	66		67
Third order "effective," decibels	60		60.5
Equivalent noise resistance, ohms	460	180	

* Ratio of product to fundamental at output for a 0.1-volt rms signal from grid to cathode.

switch to the protection line. At first sight, it would appear that using parallel tubes in each stage should greatly decrease the probability of line failure. A study of *L1* experience, however, showed that most tube failures could either be forestalled by preventive maintenance, or else were of such a nature (for example, shorts within the tube) that the parallel tube would not afford protection against line failure. The reliability advantage of parallel tubes, then, turns out to be small; their use, on the other hand, increases the wiring complexity greatly when sockets are used. The small gain in reliability would not compensate for the degradation in performance caused by the complication of wiring and the resulting capacity penalties. The applied voltages and current drains are shown in Table II.

Heat Dissipation

In common with many modern designs with high-gain vacuum tubes, the problem of heat dissipation was acute in the *L3* amplifier. The amplifier is enclosed in a sealed housing of sufficient size to dissipate readily the heat generated. However, with the usual types of chassis and tube-shield construction, vacuum-tube envelope temperatures were so high that long tube life could not be assured. Consequently a new type of tube shield was developed. This shield is of heavy copper tubing equipped with internal helical springs mounted in such a manner that each turn of the spring makes contact with both the glass tube envelope and with the copper tube, as shown in Fig. 6. Thus, each turn of the spring provides two metallic heat conducting paths to carry away the heat from the tube envelope. A total of 480 conducting paths are provided for each of the large tubes used in the output amplifier by the use of these springs. Good contact is made between the copper tubes and the chassis, so that heat generated within the vacuum tube is efficiently conducted through metallic paths to the copper chassis of the amplifier. In order not to raise the temperature of the chassis and consequently the temperature of the circuit components, large ribs are provided in the housing, with which the chassis makes intimate contact. A continuous metallic conducting path is thus provided from vacuum tube envelope to amplifier housing. This design resulted in a temperature drop in the output amplifier tubes of about 70 degrees centigrade without any temperature rise in the chassis or circuit elements over that produced when ordinary types of shields were used. The smaller vacuum tubes in the input amplifier do not ordinarily run as hot as those in the output amplifier, and a temperature drop of only about 55 degrees centigrade was attained by these methods.

Coupling Networks

The input and output coupling networks are essentially identical. The low side of each transformer is a balanced center-tapped winding which, together with the balancing network, acts as a hybrid, to produce a good 75-ohm impedance facing the cable. The use of this type of connection gives a signal-to-noise advantage over a brute-force high-side terminating network. The advantage is theoretically 3 db in the case of the output coupling network, and would approach the same figure in the case of the input network if tube noise were dominant over the resistance noise of the cable.[4] Aside from the fact that the design of a high-side shunt termination network for an off-ground peaked transformer is well-nigh impossible, the use of a balancing network in a hybrid connection has the important additional advantage that the adjustment of this network to obtain a good reflection coefficient has negligible effect on the insertion gain of the circuit.

A relatively modest share of the total shaping required has been allocated to the coupling networks: 5.5 db each, or 11 db total. One reason for this is that although these networks are outside the feedback loops in the usual sense, nevertheless the impedances which they present to the amplifiers are important factors in the feedback design, and the effects which they produce must not be allowed to become so severe as to limit the feedback to too low a value. It is obvious from inspection of Fig. 1 that only a part of the voltage developed across the input beta circuit by the plate current of the second tube will appear as a grid-cathode voltage to drive the first stage. The proportion of the beta circuit voltage which will be thus effective in producing feedback around the loop will be dependent on the potentiometer division between the impedance of the coupling network and the grid-cathode impedance of the first tube. The greater the peaking of the input coupling network, the greater its impedance at high frequencies, where the grid-cathode capacitive impedance is already decreasing, and hence the greater the potentiometer term loss. A similar loss occurs in the output amplifier. The plate current of the output stage divides between the output coupling network and the parasitic admittances to ground. The portion of the plate current which returns directly to cathode (ground) through the latter path, without passing through the beta circuit, is not effective in producing loop feedback.

A second, and even more important, limitation is that the sensitivity of the insertion gain to variations in the cou-

Table II. Power Requirements of Amplifier Unit

Applied Voltage	Current Drain
6.3 volts, grounded	1.5 amperes
6.3 volts, 190 volts off ground	0.45 ampere
+100 volts, regulated	0.5 milliampere
+190 volts	68 milliamperes
+315 volts	41 milliamperes

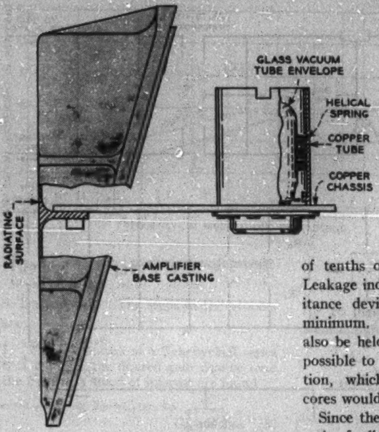


Fig. 6. Tube shield, for heat conduction

pling-network elements is increased as the slope is increased. The same considerations lead us to keep not only the slope but also the gain level or efficiency of these networks relatively low. The maximum-possible coupling network gain which can be obtained over the entire frequency spectrum is limited by the capacity across the circuit, as shown by Bode's resistance integral theorem. This capacitance cannot be reduced without incurring a more severe potentiometer term penalty and thus limiting feedback, so the total gain area cannot be profitably increased. The in-band gain can be made greater or less, as we concentrate more or less of the total gain area in the transmitted band, but it is found that attempting to get high values of in-band gain area leads to networks which are increasingly sensitive to element deviations. A resistance area efficiency of about 50 per cent turns out to be the most acceptable compromise.

In spite of these steps, the coupling networks are the most important source of manufacturing deviations, but by improved mechanical design and the use of quality control, these effects have been reduced by an order of magnitude as compared to previous designs. For example, the end capacitance of each coupling network is a quartz-disc capacitor, and the peaking or splitting coils are tension-wound on ceramic forms to give inductances of 1 per cent tolerance with distribution requirements within this range. The windings of the transformers, shown in Fig. 7, are made by plating the turns on threaded forms machined from optical-grade quartz or Vycor glass, to tolerances of tenths of a thousandth of an inch. Leakage inductance and parasitic capacitance deviations are thus held to a minimum. Split ferrite cores, which must also be held to close tolerances, make it possible to use these methods of fabrication, which previously available tape cores would not have permitted.

Since the amplifier configuration gives series feedback on the tubes adjacent to the cable, and cathode feedback on the tubes adjacent to the regulating network, the high-impedance winding of each coupling network is necessarily off-ground. This leads to considerable difficulty in specifying the equivalent circuit. For an on-ground coupling network, the equivalent circuit of Fig. 8(B) would be adequate for gain and feedback computations—essentially a reactive equalizer plus an ideal transformer. Even then, the capacities and inductances associated with the transformer itself would have to be given as functions of frequency because of the distributed nature of the device. When, however, the high-impedance winding is raised above ground potential by the voltage developed across the beta circuit, which is nearly the same magnitude as the voltage across the coupling network, the effects of distributed parasitic capacitances to ground, and the lumped capacitance from the junction of transformer and peaking coil, become of prime importance. The coupling network, therefore, cannot be adequately represented by merely lifting the circuit of Fig. 8(B) off-ground.

In order correctly to understand and compute the amplifier gain, it was necessary to develop a complex mathematical analysis of the distributed structure of the transformer, in conjunction with an extended program of precise measurement of the transformer constants. Even then, one must be content with an accuracy of a few tenths of a decibel and a few degrees of phase, as compared with the order-of-magnitude better accuracy which can be obtained for a 2-terminal lumped-constant network. The agreement between measurement and computation of amplifier gain and reflection coefficient is sufficiently good, however, to assure that all the important effects are sufficiently understood to make intelligent control possible.

Using the results of this analysis, the off-ground coupling network can be represented by the equivalent circuit of Fig. 8(C), where the values of the pi of admittances are obtained in terms of the fundamental parameters of the transformer and associated elements. This representation is convenient for insertion gain and feedback computations. The most important difference between the equivalent circuits of Figs. 8(B) and 8(C), from a practical standpoint, is the presence in the latter of the admittance Y_{1CN}, which is a manifestation of the capacitance to ground from the high impedance winding and the junction of transformer and peaking coil. I_2 and the contributions to Y_{2CN} not directly attributable to the obvious shield-to-shield capacity of the transformer can be set equal to zero with little error, but Y_{1CN} affects the insertion gain and feedback by about 2 db.

It will be observed that the transformer is shown with two shields, one connected to the bottom of the high-impedance winding, which is the top of the beta circuit; the other connected to ground. The first of these, which is physically adjacent to the high-impedance winding, acts to collect, and carry to the beta circuit, the distributed capacitance of the high winding, thus avoiding intolerably large capacitance from this winding to ground. The second shield prevents capacitative coupling of the large beta-circuit signal voltage to the low-impedance winding, which would lead to a very poor reflection-coefficient performance. Typical curves of the reflection coefficient are shown in Fig. 9. Since the amplifier is an active device, the reflection coefficient is to some extent a function of the vacuum-tube transconductances, and tends to be degraded as tubes age.

Amplifier Equations

Using the equivalent circuit of the coupling network shown in Fig. 8(C), the circuit of the input amplifier can be represented as in Fig. 10, and the equations shown can be derived from straight forward nodal analysis of the circuit. Similar equations can be derived for the more complicated output amplifier. From these, using Thevenin's theorem, the gain of the tandem combination can be computed, as well as the feedbacks on the various tubes.

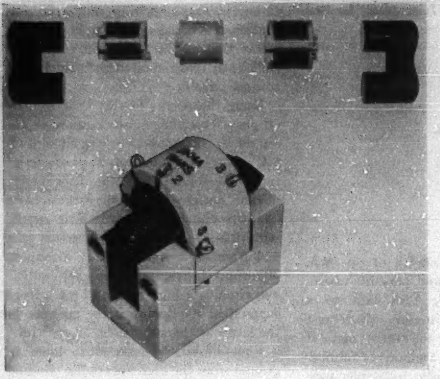


Fig. 7 (above). The precision transformer: windings, core, and assembled unit

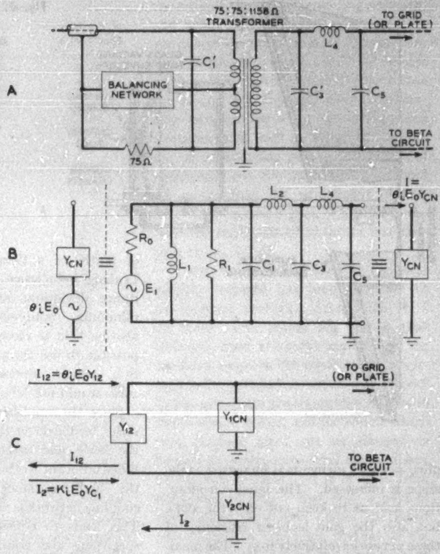


Fig. 8. Coupling network

A. Coupling-network elements
C_1'—Low side padding capacitance
C_3'—High side padding capacitance
L_4—Peaking coil
C_5—Peaking capacitance

B. On-ground equivalent circuit
E_1—Equivalent generator, $\sqrt{1158/150}\times$ cable open circuit voltage, E_0
R_0—1,158 ohms
L_1, R_1—Mutual inductance, dissipation of transformer
C_1—Low side capacity$\times\sqrt{150/1158}$
L_2—Leakage, referred to high side of transformer
C_3—High side capacity of transformer (including padding capacitor)
L_4, C_5—Peaking elements

C. Off-ground equivalent circuit
Note: Where Y_{C_1} in I_2 formula is admittance of cable

Similarly the input amplifier can be replaced, for convenience in reflection-coefficient calculations, by the pi of admittances and the driving current of Fig. 10(B). The equations of Fig. 10(B) expressed in terms of the cofactors of the circuit determinant are of general application for the reduction of a multinode circuit to simpler form.

Regulating Network

Like the coupling networks, the regulating network between the amplifiers is outside the feedback loops, and the gain of the amplifier is very nearly a direct function of the impedance seen looking into the network. This impedance is controlled by a single variable resistance—the thermistor—which is directly heated by the d-c output current of the regulator. The output of the regulator, in turn, is a function of the magnitude of main pilot level at the output of the amplifier. The whole forms an automatic-volume-control circuit, which acts to keep the level of the main pilot constant at amplifier output, under the assumption that any changes in this pilot level are caused by line temperature or length deviations. The invariant arms of the regulating network are designed to give the accurate shaping of network impedance versus frequency required so that the amplifier gain change will have the wanted square-root of frequency shape. The design of the network is fundamentally based on the variable-equalizer theory developed by H. W. Bode,[5] but the process of finding physical networks which will give the desired performance to a very high order of accuracy makes use of newer methods developed by S. Darlington.[6,7] Since these methods were also used in the design of several other amplifier networks, and because the precision thus made possible is essential to the system, a brief recapitulation of the steps involved is of interest.

Fig. 11 shows the basic configuration of the regulating network. It was selected because it is one of the simplest that gives symmetrical gain changes controlled by a single resistor. It is also capable of absorbing the parasitic interamplifier capacitance, and gives some advantageous gain shaping at normal setting, although this shaping is not under design control.

We need to find the impedance of the two arms Z_1 and Z_2 that will give the required square-root of frequency shape of gain change (with a value of 6 db at 8 megacycles) as the thermistor is varied by a factor of three. For this circuit, the change in gain may be expressed as a ratio of two impedances

$$\epsilon^{\alpha+j\beta}=\frac{\mathbf{Z}_H}{\mathbf{Z}_N}=\frac{\dfrac{1}{\mathbf{Z}_1}+\dfrac{1}{\mathbf{Z}_2+R}}{\dfrac{1}{\mathbf{Z}_1}+\dfrac{1}{\mathbf{Z}_2+kR}}=F(p) \tag{1}$$

$$\epsilon^{-\alpha-j\beta}=\frac{\mathbf{Z}_L}{\mathbf{Z}_N}=\frac{\dfrac{1}{\mathbf{Z}_1}+\dfrac{1}{\mathbf{Z}_2+R}}{\dfrac{1}{\mathbf{Z}_1}+\dfrac{1}{\mathbf{Z}_2+\dfrac{R}{k}}}=\frac{1}{F(p)} \tag{2}$$

where

$\mathbf{Z}_N$ = the imepdance looking into the network with a thermistor value R

$\mathbf{Z}_H$ = the impedance looking into the network with a thermistor value kR

$\mathbf{Z}_L$ = the impedance looking into the network with a thermistor value R/k

For this design, $R = 500$ ohms and $k = 3$

We may solve equations 1 and 2 and obtain expressions for $\mathbf{Z}_1$ and $\mathbf{Z}_2$. However, for synthesis $\mathbf{Z}_1$ and $\mathbf{Z}_2$ should be expressed in terms of singularities in the p-plane. This may be accomplished by approximating the wanted gain change with a Tchebycheff series, as developed by

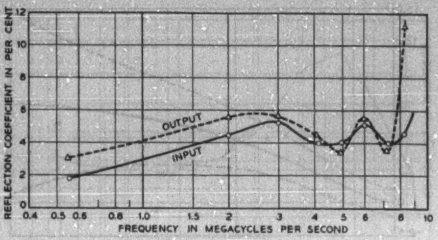


Fig. 9. Reflection coefficient performance

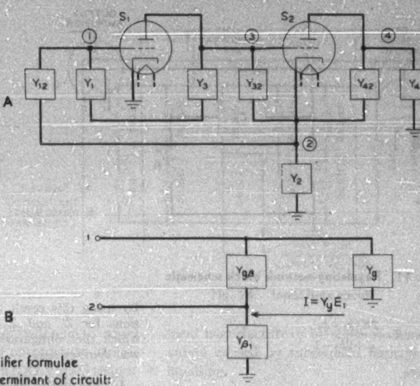


Fig. 10 (right). Input amplifier

S. Darlington.[7] The steps in the process are:

1. The coefficients of a Tchebycheff series that matches the desired gain change over the frequency band of interest are found

$$\alpha = C_0 + C_2 \cos 2\phi + C_4 \cos 4\phi \ldots C_{2n} \cos 2n\phi \quad (3)$$

where the relation between ω and ϕ is given by the band-pass transformation

$$\omega^2 = K\frac{k_1 + \sin^2\phi}{k_2 - \sin^2\phi}$$

$$K = \omega_{c1}\omega_{c2}$$

$$\omega_{c1} = K\frac{k_1}{k_2} = \text{lower cutoff frequency, 0.2 megacycle}$$

$$\omega_{c2} = K\frac{k_1+1}{k_2-1} = \text{upper cutoff frequency, 8.35 megacycles}$$

and

$$C_{2k} = \frac{2}{n+1}\sum_{r=1}^{r=n+1} \alpha_r \cos 2k\phi_r \quad \begin{matrix} k = 1, 2 \ldots n \\ r = 1, 2 \ldots n+1 \\ \text{(Match points)} \end{matrix}$$

$$\phi_r = \frac{(2r-1)}{(2r+1)}\frac{\pi}{2}, \quad \alpha_r = \text{gain at } \phi_r$$

2. The coefficients of a related power series are determined

$$\epsilon^{2\alpha} = 1 + A_2 Z^2 + A_4 Z^4 + \ldots A_{2n} Z^{2n} \equiv F(Z^2) \quad (4)$$

where

$$A_2 + S_2 = 0 \quad \text{and} \quad S_{2k} = kC_{2k}$$
$$2A_4 + A_2 S_2 + S_4 = 0$$
$$nA_{2n} + A_{2n-2} + \ldots = 0$$

3. $F(Z^2)$ is expressed as a rational fraction containing both natural modes and infinite loss points

$$F(Z^2) = \frac{N(Z^2)}{D(Z^2)} \quad (5)$$

where the coefficients of N and D may be found from the continued fraction expansion of $F(Z^2)$. The degree of N and D fixed by the allowable approximation error and the complexity of the network.

4. The roots of N and D (in terms of Z^2) are found and then transformed to the p-plane using the transformation

A. Input amplifier formulae
Nodal determinant of circuit:

$$\Delta = \begin{array}{cccc|c} E_1 & E_2 & E_3 & E_4 & \\ \hline Y_{11} & -Y_{12} & 0 & 0 & I_1 \\ -Y_{12} & Y_{22}+S_2 & -Y_{32}-S_2 & -Y_{42} & I_2 \\ S_1 & -Y_{32} & Y_{33} & 0 & I_3 \\ 0 & -Y_{42}-S_2 & S_2 & Y_4+Y_{42} & I_4 \end{array}$$

where $Y_{11} = Y_1 + Y_{12}$; $Y_{22} = Y_2 + Y_{12} + Y_{32} + Y_{42}$; $Y_{33} = Y_3 + Y_{32}$

Feedbacks:

$$\text{Return ratio on first tube: } T_1 = \frac{P_{12}\frac{S_1S_2}{Y_{33}Y_2'}P_4 + P_{12}\frac{S_1}{Y_2'}P_{32}}{1 + P_3\frac{S_2}{Y_2'}P_4} = \frac{B+D}{1+E}$$

$$\text{on second tube: } T_2 = \frac{P_{12}\frac{S_1S_2}{Y_{33}Y_2'}P_4 + P_3\frac{S_2}{Y_2'}P_4}{1 + P_{12}\frac{S_1}{Y_2'}P_{32}} = \frac{B+E}{1+D}$$

$$\text{where } P_{12} = \frac{Y_{12}}{Y_{11}}, \; P_4 = \frac{Y_4}{Y_4 + Y_{42}}, \; P_3 = \frac{Y_3}{Y_3}, \; P_{32} = \frac{Y_{32}}{Y_{33}}$$

$$Y_2' = Y_2 + \sigma_1 + \sigma_3 + \sigma_4; \; \sigma_1 = \frac{Y_{12}Y_1}{Y_{11}}, \; \sigma_3 = \frac{Y_3Y_{32}}{Y_{33}}, \; \sigma_4 = \frac{Y_4Y_{42}}{Y_4 + Y_{4_2}}$$

Transmission:

$$E_4 = I_{12}\frac{\Delta_{14} - \Delta_{24}}{\Delta} + I_2\frac{\Delta_{24}}{\Delta} \text{ using for } I_{12} \text{ and } I_2 \text{ the values given on Fig. 8(C)}$$

$$E_4 = E_{0e}\left[\theta i\frac{Y_2}{Y_4}\,\frac{B - D\frac{\sigma_4}{Y_2} - E\frac{\sigma_1}{Y_2} - \frac{\sigma_1}{Y_2}\frac{\sigma_4}{Y_2'}}{1+B+D+E} + KiY_{C1}\frac{1}{Y_4}\,\frac{B+E+\frac{\sigma_4}{Y_2'}}{1+B+D+E}\right]$$

B. Equivalent circuit of input amplifier as seen by coupling network

$$Y_g = \frac{\Delta_{22} - \Delta_{21}}{\Delta_{12-12}} = Y_1 \qquad Y_{\beta 1} = \frac{\Delta_{11} - \Delta_{21}}{\Delta_{12-12}} = Y_2 + \sigma_3 + \sigma_4 + P_3S_2P_4$$

$$Y_{g\beta} = \frac{\Delta_{21}}{\Delta_{12-12}} = Y_{12} \qquad Y_y = \frac{\Delta_{12} - \Delta_{21}}{\Delta_{12-12}} = -\left[P_{32}S_1 + \frac{S_1S_2}{Y_{33}}P_4\right]$$

where the cofactor Δ_{12-12} is found by striking out the first and second row and first and second column of the circuit determinant, the sign following the usual rules for the sign of a cofactor

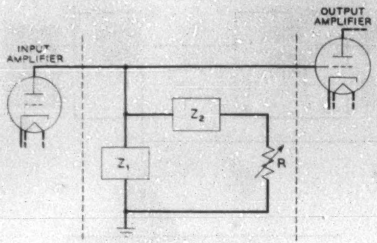


Fig. 11. **Regulating-network block schematic**

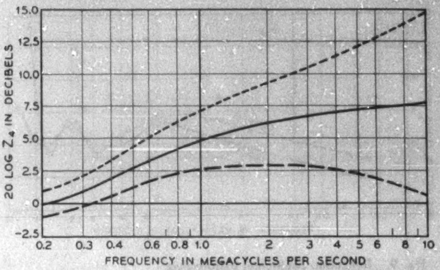


Fig. 12. **Regulating-network gain versus frequency—for extreme and mid-range thermal settings**

$$p^2\sigma = K - \frac{2K(k_1+k_2)}{\psi+2k_2-1}$$

where

$$\psi = \frac{1+\boldsymbol{Z}^4}{2\boldsymbol{Z}^2} \tag{6}$$

These four steps result in a polynomial $F(p)$ satisfying the requirements on the change in gain. In this specific case

$$F(p) = \frac{1.06(p+0.0960)(p+0.0280)(p+0.9890)(p+0.2890)}{(p+0.1058)(p+1.803)(p+0.0300)(p+0.3492)} \cong \epsilon^{\alpha+j\beta} \tag{7}$$

Solving equations 1 and 2 obtain

$$\boldsymbol{Z}_2 = \frac{k-F}{kF-1}R, \qquad \boldsymbol{Z}_1 = \frac{(k^2-1)(F^2-1)}{(kF-1)(k-F)} \tag{8}$$

where

$$F(p) \equiv F$$

Since the design must absorb the interstage capacitances (and also the stray capacitances of some of the elements in the physical network), $\boldsymbol{Z}_1$ must include a shunt capacitor. It can be shown that the desired result is obtained when

$$F \rightarrow 1 + g\frac{1}{p} \text{ as } p \rightarrow \infty \tag{9}$$

Since the derived transmission function F, equation 7, does not satisfy the condition expressed by equation 9, the function must be modified in order to provide a shunt capacitor in $\boldsymbol{Z}_1$. To accomplish this we use equation 5 expressed as a bilinear form of the *n*th term of the continued fraction expansion.

$$F(\boldsymbol{Z}^2) = \frac{N_1(\boldsymbol{Z}^2)+K_nN_2(\boldsymbol{Z}^2)}{D_1(\boldsymbol{Z}^2)+K_nD_2(N^2)} \tag{10}$$

The coefficient K_n is modified so that the new $F(p)$ satisfies equation 9. The modification of K_n does not substantially alter the required in-band transmission and also does not produce nonphysical singularities. The new F obtained is

$$F(p) = \frac{(p+0.0289)(p+0.1018)(p+0.3431)(p+1.017\pm j1.628)}{(p+0.0310)(p+0.1132)(p+0.4426)(p+0.6158\pm j1.368)} \tag{11}$$

By using this result in equation 8 expressions for $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ were obtained from which the configuration shown in Fig. 11 was synthesized.

Fig. 12 shows the voltage which would be developed across the regulating network versus frequency in response to a constant current driving force, for the midrange and extreme values of thermistor resistance. The slope across the band for the midrange value is 7.5 db; this is the regulating network contribution to the total slope of 37 db required of the complete amplifier.

To the extent that the differences between the curves of Fig. 12 fail to exactly match the desired square-root of frequency characteristic, the action of the regulating network will introduce an equalization error. This regulation error is shown in Fig. 13. It amounts to a few hundredths of a decibel for a 6-db gain change, caused in part by network design imperfections and in part by the fact that very small second-order effects result in the amplifier gain not exactly following the regulating-network impedance.

Beta Circuits

Starting from our basic requirement that a slope of 37 db across the transmitted band must be obtained, and noting that the total of the contributions from the coupling networks and regulating network is 16 db, we are left with about 20 db of shaping to be supplied by the beta circuits. The input beta circuit has been designed to supply most of this remainder, the contributions of the output beta circuit and the $\mu\beta/1-\mu\beta$ effects in amplifiers being only 3 db. The original design procedure for this network was basically the same as for the regulating network, but the final values of the elements represent a long process of incorporating secondary corrections as our knowledge of the amplifier grew. Original constraints included the necessity of including as part of the beta circuit not only the relatively large parasitic capacitance of the amplifier and the coupling network shield to shield capacitance, but also the screen resistor and by-pass capacitor of the second tube, which, as usual, is by-passed not to ground but to cathode, so that modulation product components of the screen current will not be introduced into the signal path via the beta circuit. It is also required that the beta circuit have the correct d-c resistance to serve as the cathode-bias resistor of the second tube, and that it incorporate provision for metering this bias through suitable decoupling elements, which are among the gain-determining elements of the network. Finally, this network was used as the mop-up equalization network of the amplifier, and its element values were readjusted to give the correct amplifier gain after the performance of a representative group of amplifiers with average coupling networks and tubes had been determined. In doing this tailoring, it is necessary to precorrect for the effects of noninfinite feedback, since the gain of the amplifier is not exactly a direct function of the beta-circuit admittance. The configuration of the input beta circuit is shown by Fig. 14; it is a simple 2-terminal network. Its in-band impedance varies from about 1,000 ohms at 300 kc to 110 ohms at 8.35 mc.

The output beta circuit is relatively flat, which in this case is the optimum condition for signal-to-noise and feedback-loop stability considerations. Because of this simplicity, it is possible to incorporate in this network provision for adjusting the gain of the amplifier, to reduce the misalignment of the system.

Misalignment Adjustment

We can distinguish three major effects which will contribute to misalignment

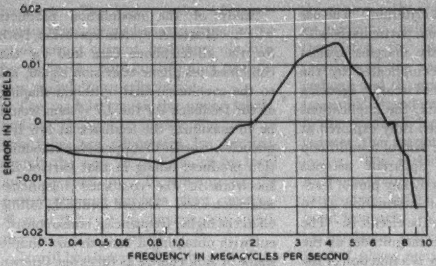


Fig. 13. Regulation error per 1-db regulation

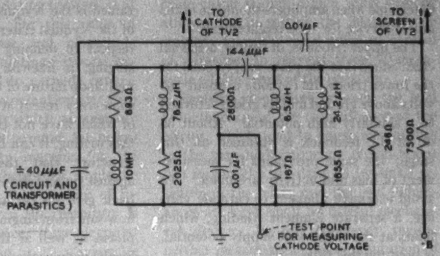


Fig. 14. Input beta circuit

and consequent degradation of system signal-to-noise ratio:

1. Design error—the degree to which the design gain of the amplifier, because of the finite number of elements and other limitations, fails to match the line loss.

2. The cumulative effect in the individual amplifier of manufacturing deviations of the elements.

3. The aging of the components of the amplifier, of which the tube aging will, of course, be the dominant short-term effect.

The signal-to-noise performance of the system can be improved by reducing the misalignment, if this is done without resorting to measures which would introduce systematic instead of random gain deviations.

If we study the shapes of gain change introduced by the more important deviation contributors, particularly the element variations, we find, as might be expected, that in general the effects are small at low frequencies and increase sharply near the upper edge of the band, if the thermistor is held constant. In the system, the regulation around the main pilot will automatically act to reduce the deviation at 7 megacycles to zero, adding a square root of frequency curve that results in a bow-shaped deviation, as illustrated by Fig. 15. Examining the signal-to-noise effects of the degradation caused by misalignment, we conclude that it is most desirable to reduce as much as possible the misalignment at 4 megacycles, the television carrier frequency in the combined system. The output beta circuit has, therefore, been designed to give varying amounts of this shape, the total range being ±0.6 db at 4 megacycles, in 0.2-db steps. The successive steps of this gain adjustment are simple multiples of each other, symmetrical in the two directions of adjustment, so that we put into the system a single systematic shape, which will be small in magnitude at the equalizing points because the different settings of this control in successive line repeaters will tend to cancel out. This setting of output beta circuit gain is controlled by a *GAIN ADJ* switch, accessible from outside the amplifier housing. The adjustment will be made on an in-service basis as part of normal maintenance procedures, using the level at each repeater of one of the system pilot frequencies, 3,096 kc, as the index of proper setting.

Manufacturing Testing

As mentioned above, the mechanical design of the amplifier has been planned to permit the separate testing of the five gain-determining networks, the tuned interstage of the input amplifier, and the separate input and output amplifiers, before their assembly with the regulating network to form complete amplifiers. Variations in environment are minimized by the use of jigs which also make it unnecessary, in general, to solder to the network under test. Visual gain sets which cover the transmitted band and are accurate to 0.02 or 0.03 db are used. The component network or component amplifier under test is connected in series with a complementary or equalizing network, to obtain a flat transmission characteristic which can be accurately compared to the transmission of standard attenuators. The completely assembled amplifiers are similarly tested. Quality-control charts of the resulting measurements on all components are useful in detecting shifts in transmission which might be caused by loss of control in element manufacture or by shifts of element values caused by subsequent damage in handling.

D-C Feedback

In addition to the loop feedback at signal frequencies, local d-c feedback is used on each stage. The grid of each tube is returned to a +9-volt potential rather than to ground, and about +10 volts is developed across each cathode resistor. The usual stabilizing effects of self-bias are thus obtained in exaggerated degree, each tube having about 20 db of local d-c feedback. Care must then be taken to select the cathode by-pass and interstage coupling condensers so that the transition from low-frequency local feedback to in-band loop feedback is accomplished smoothly, without the instability which might be caused by a balancing out of these two feedbacks in the transition region.

For maintenance measurements, the 9-volt bias potential and the d-c cathode voltage of each tube are brought out to a multipin amplifier test jack through appropriate decoupling filters. Thus, the bias on each stage can be measured on an in-service basis. Filament-voltage dropping resistors which can be switched in or out of circuit are provided on the repeater panel, so that bias can also be observed for a filament supply voltage 10 per cent below normal. The activity or change in plate current with filament voltage thus measured, or the history of the bias at normal filament voltage, will be used to

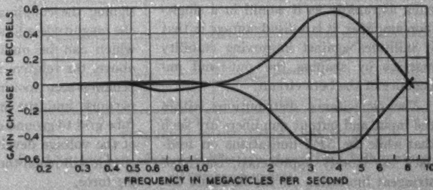


Fig. 15. Misalignment adjustment

determine when amplifiers should be taken out of service for tube replacement.

The upper triode is, of course, a special case: since it is the plate-supply path for the lower triode, its cathode is about 160 volts above ground and its grid is returned to a similarly high potential. About 35 db of d-c feedback is obtained on this stage; the same provisions for measurement of bias and activity are made. To avoid excessive filament to cathode voltage, a separate filament winding, which floats at the +190-volt supply potential, is used for this tube.

Loop Feedback

The design of the feedback loops of the input amplifier follows conventional practice. The constraints which operate to limit the amount of feedback which can be obtained in the transmitted band are well known.[8] Broadly speaking, the figure-of-merit of the vacuum tubes and the circuit capacitances determine the asymptotic cutoff, which in any feedback amplifier limits the magnitude of the feedback which can be built up in the band. In multiloop structures, there are additional limitations. Consider, for example the formula given in Fig. 10(A) for the feedback on the second tube. If we increase without limit the magnitude of the first tube transconductance, we find that the feedback on the second tube approaches the value $S_1P_4Z_{32}$. This is a limit, common to multiloop structures, over and above the usual limit imposed by the Nyquist stability criterion. A similar *SPZ* limit can be derived for the feedback on each stage. The same multiloop mechanisms which operate to limit the feedback also result in making the feedback on any given stage relatively insensitive to variations in other transconductances or impedances. In a single-loop circuit, a 1-db change in beta-circuit impedance or first-stage transconductance causes a 1-db change in the feedback on every tube. In the multiloop case, the feedback on, say, the second stage would generally change only 1/2 db at most frequencies. While this is an advantage, in the sense that transconductance decay does not decrease feedback as rapidly as it would in a single-loop amplifier, it is a disadvantage in that it militates against improving stability margins by shaping the out-band impedance of the beta circuit.

The capacitance distributions within the input and output amplifiers are such that while the *SPZ* limitations on feedback are closely approached, the most stringent limitation on the feedback obtained is the Nyquist criterion. The use of the Nyquist criterion, particularly with respect to defining the margins against singing, is likewise complicated by the multiloop nature of the circuit. Ignoring some very recent work, the implications of which have not been fully explored at this writing, it can be said of a multiloop structure that the apparent margins against singing shown by any plot of feedback give no certain information as to how safe from singing the circuit is. The phase, as well as the magnitude, of the feedback on each stage is a function of the magnitude of the other transconductances and either decay or increase of these transconductances might destroy the phase margin. In these circumstances, it is theoretically necessary to examine for stability every conceivable combination of transconductances. A more practical expedient, of course, is to rely on judgment, backed by computation and laboratory experiment on the circuit for a wide, but far from infinite, number of circuit conditions.

Another difficulty arises from the fact that the feedback on each tube is different, so that gain and phase margins obtained for one return ratio do not imply equal gain and phase margins for the return ratio on some other tube of the same amplifier. It does not follow, however, that it is necessary to investigate separately the margins on each stage versus circuit element variations. The point to be stressed here is that we are using the behaviour of the return ratio merely as an index of our real concern: the position on the p-plane of zeros of the determinant of the circuit, and the determinant is the same for both stages. Rather than asking how many decibels of gain margin and how many degrees of phase margin any particular return ratio displays, we ought instead to inquire how quickly the apparent margins disappear as we change transconductances and network impedances. The margins on all the return ratios will vanish simultaneously, regardless of the apparent difference in original margin magnitudes. It is therefore satisfactory to examine the behaviour of whichever return ratio is most easily observed.

The choices made in the course of designing the coupling network also affect the magnitude and shape of the feedback which can be obtained: as mentioned above, the relative magnitude of the impedance seen looking into the coupling network and the impedance from first tube grid to ground determines how much of the voltage developed across the beta circuit will reach the first stage as a driving force.

Study of the modulation products which will arise in system operation, both for the all-telephone case and for the combined telephone-television signal, led to the conclusion that optimum shaping of the feedback for the *L3* system would be to maximize the feedback at low frequencies, in order to suppress intermodulation products falling in that part of the spectrum in the combined telephone-television case. Shaped feedback, falling off at the higher frequencies, is also consistent with obtaining a smooth and simple shape of gain-change as tubes age (known as "mu-beta effect"), which is desirable from the equalization standpoint. With these considerations in mind, the interstage of the input amplifier has been peaked well above the transmitted band (at 11 megacycles) to partially compensate for the input potentiometer term, and to help in achieving this smooth shape of mu-beta effect. If flat feedback over the band were the objective, it would also be necessary to design the grid-cathode admittance of the second tube so that the parasitic grid-cathode capacity would be absorbed in a flat impedance versus frequency, but in this case the desired shaping of the second-tube feedback is attained by taking advantage of the way in which the grid-cathode capacitance naturally limits the high-frequency feedback on this stage.

The loop feedback in the output amplifier is similarly shaped, for the same modulation and equalization reasons. The use of the double-triode circuit in this amplifier is an unusual feature. This connection of two triodes, sometimes referred to as the "cascode circuit," has appeared in many contexts in recent years, usually to serve some other purpose than here. It serves as a superior output stage in the *L3* amplifier largely because the effective transconductance which can be obtained is about 3-db higher than that of a pentode of the same grid-cathode spacing. The effective transconductance, ignoring for the moment the division of output current between the load impedance and the internal plate impedance of the upper triode, is very nearly the transconductance of the lower triode. Since no current is lost to a screen, this is higher than that of a pentode of the same spacing. The output impedance of the upper triode is multiplied by the local feedback consequent on its cathode being off ground by the plate impedance of the lower triode. Since the load presented to the lower triode is a low impedance, approaching the reciprocal of the transconductance of the upper triode, the grid-plate capacitance of the lower triode is not enhanced by feedback. The higher value of transconductance means less grid drive for the same output current, and more obtainable feedback; both contributing to superior modulation performance.

The modulation of the upper triode, and the changes in transconductance of this tube, are suppressed by approximately the sum, in decibels, of the local and loop feedback. In consequence, the modulation and mu-beta effect contributions of this tube to the complete amplifier are negligibly small.

It will be observed that the grid of the upper triode is not connected to ground, but to the top of the beta circuit. In consequence, the grid-plate capacitance of the upper triode appears as part of the end capacitance of the coupling network, rather than as a parasitic capacity to ground. This results in a gentle potentiometer term in the output amplifier. This connection, however, also has the effect of vitiating to some extent the desirable qualities of the circuit, particularly at the higher frequencies, where the grid-cathode capacitance of the upper triode becomes important.

Because of this gentle output-potentiometer term, it is not necessary to tune the interstage of the output amplifier, which consists simply of the circuit capacitance plus a network which has the characteristics of a 10-micromicrofarad capacitance in the transmitted band. Above the band, this network shapes the gain and phase characteristics of the feedback loop to obtain the desired stability margins. The incorporation of this network reduces the in-band feedback by about 3 db, a sacrifice which unfortunately is necessary to assure stability when the thermistor in the regulating network is at its minimum value. For this value of thermistor, the phase and magnitude relations of the regulating-network impedance and the grid to cathode capacitance of *VT3* produce a potentiometer term in the feedback loop which appreciably reduces the margins around 30 megacycles. The stability margins for the midrange value of thermistor would be satisfactory without this sacrifice of in-band feedback. When the thermistor is at maximum resistance, some degradation of phase margin at 10 megacycles occurs, again because of the potentiometer term effect mentioned in the foregoing, but in this case the remaining margin is sufficient, since the circuit elements are still under good control at this frequency. Because the plate-cathode impedance of *VT2* is very high, the similar potentiometer term at the output of the input amplifier causes only negligible changes in input amplifier stability margins as the thermistor changes.

In the 70-megacycle region, there are two almost equally important feedback loops in the output amplifier—one through the transconductance of the lower triode; the other through the grid-plate capacitance of this tube. Balances between these feedback paths are observed in the 70 to 100-megacycle region in the course of measuring the feedback, sometimes accompanied by 180-degree shifts in the phase of the loop transmission at frequencies above the balance point, an effect which theoretically depends on just how the two vectors go through the balance point. The occurrence of these balances is accompanied by a few degrees loss of phase margin in the 30-megacycle cutoff region, which must also be allowed for in setting the 30-megacycle stability margins, since sufficient control of parasitics to prevent these 70-megacycle effects is out of the question.

Parasitic resonances between the lead inductances and the capacitances of the circuit, which tend to cause instabilities in the very-high-frequency region about 200 megacycles are damped by small resistors in the leads, and the lead inductances are kept small by careful mechanical design. In this frequency region, neither measurement nor computation of stability margins can be trusted as anything but a rough guide. On the other hand, adding damping resistors in grid leads and other critical points to prevent 200-megacycle sings causes a phase margin penalty in the 30-megacycle region, so a nice judgment of how much damping to add is called for. Final values of damping were chosen so that typical amplifiers could not be made to sing by increasing critical lead lengths, or by substantial increases in parasitic capacity, thus assuring that manufacturing variations of elements and wiring will not cause high-frequency sings. The return ratio of *VT4* is shown on Fig. 16, the mu-beta effects of both amplifiers on Fig. 17.

Signal Levels, Modulation, and Noise

Fig. 18 shows the signal levels within the amplifier in decibels relative to one volt from grid to cathode of *VT4*, which is a convenient point to use as a reference for system signal-to-noise studies. It will be noted that as a result of using the input beta circuit to give so much of the shaping of amplifier gain, the input ampli-

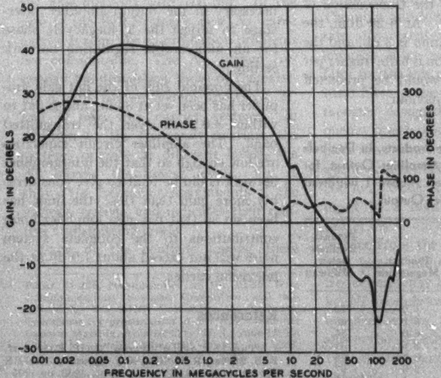


Fig. 16. Feedback on lower triode versus frequency

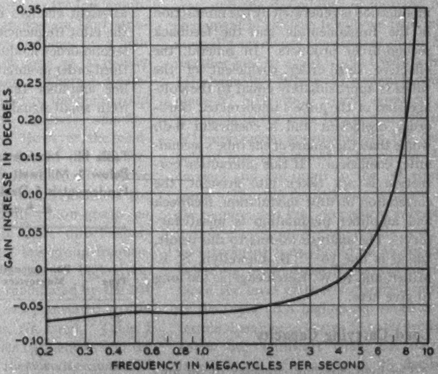


Fig. 17. Mu-beta effect—amplifier gain change for a 1-db increase in all transconductances

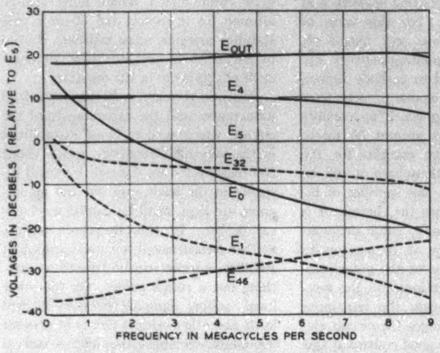


Fig. 18. Signal levels within amplifier

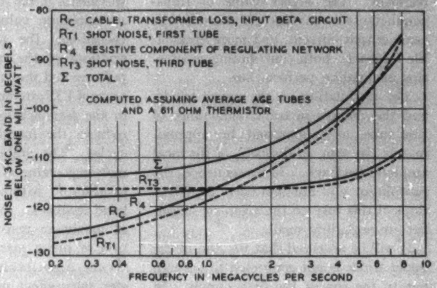


Fig. 19. Thermal noise at line amplifier output

fier has little gain at low frequencies. In consequence, the input tube of the output amplifier and the regulating network are important thermal-noise sources at low frequencies. The relative magnitudes of the noise sources are shown on Fig. 19, which gives the noise at amplifier output as a function of frequency.

Comparison of the grid to cathode voltages of *VT2* and *VT4* shows that the former will be an important modulation contributor, since the driving force on these tubes is nearly equal, particularly at low frequencies. Typical amplifier modulation values are given in Table III. Computations using the measured feedback and the performance of single tubes without feedback check the measured values of amplifier modulation to within a couple of decibels if the third-order coefficient of the tube is corrected to take account of the fact that some third-order modulation is generated by the interaction of the fundamentals and the feedback second-order products. In general, the effective third-order coefficient of the tubes is approximately equal to the voltage sum of the tube's uncorrected third-order coefficient and a coefficient 6-db worse than the square of the tube's second-order coefficient. If this interaction correction is not taken into account, the correlation of tube modulation, feedback and amplifier modulation is unsatisfactory. The analysis leading to this result, which is due to F. B. Llewellyn, S. E. Miller, and R. W. Ketchledge, is too long to give here.

Load-Carrying Capacity

The load-carrying capacity of an amplifier is difficult to define with exactness. One possible definition is the load at which the modulation coefficients of the amplifier have departed appreciably from the small-signal power-series values because of loss of feedback, as the transconductance is cut off during part of the cycle. The signal carried without serious overload, in terms of a single frequency, is practically constant in the transmitted band, as a consequence of the fact that the output voltage and the lower-triode grid voltage have nearly the same shape versus frequency characteristic, as shown in Fig. 18. The output coupling-network shaping approximately compensates for the potentiometer term division of current between the load impedance and parasitic paths to ground. Departure from the small-signal power-series behaviour just begins to be appreciable at +14 decibels referred to 1 milliwatt (dbm) of any single frequency. At +18 dbm, the d-c effects of overload show up as slight changes in the transmission of the pilot frequencies. At +26 dbm, the second-order modulation is 3 db, and the third-order modulation is 6 db, higher per line amplifier than would be predicted from small-signal behaviour.

Table III. Modulation Products, in Decibels Below 1 Milliwatt at Amplifier Output, for Fundamentals 5 Decibels Above 1 Milliwatt at Amplifier Output

Type	Fundamental, Megacycles	Product, Megacycles	Product, Decibels Below 1 Milliwatt
2F	0.5	1.0	72
	1.0	2.0	65
	3.0	6.0	55
	4.0	8.0	51.5
3F	0.25	0.75	95
	0.667	2.0	97
	2.0	6.0	87.5
	2.66	8.0	81.5

Flat Amplifier

The flat-gain amplifier, which is used as a transmitting amplifier, and to make up for equalizer loss at various points in the system, is basically the same as the line amplifier, with only the obviously necessary modifications. The input beta circuit is nearly flat, the regulating network has been replaced by a fixed-gain network (which contains a single variable element whose adjustment at the factory compensates to some extent for variations in the coupling networks), and the input coupling network has been modified so that the peaking used in the line amplifier is replaced by a drop in the high-frequency gain of this network. The output beta circuit has been modified to give flat gain control of ±1.0 db in 0.2-db steps. The interstage designs are changed to readjust the feedback so that the modulation suppression and the change in gain as tubes age will be nearly the same as in the line amplifier. Somewhat more feedback is obtained in the output amplifier, since no network is necessary in the output-amplifier interstage to adjust the 30-megacycle phase for an unfavorable regulating network setting.

The nominal gain of the flat gain amplifier has been set at 34 db, and is flat to within ±0.2 db over the transmitted band. The amplifier circuit capacities are low enough so that the interamplifier network could be built to give considerably more gain than this; the limit has been set so that flat-gain amplifier noise contributions to the complete system noise will not exceed about 1.0 db at the television carrier.

References

1. The L3 Coaxial System, C. H. Elmendorf, R. S. Ehrbar, R. H. Klie, A. J. Grossman. *AIEE Transactions*, vol. 72, pt. I, Sept. 1953, pp. 395–413.

2. L3 Coaxial System—Equalization and

Regulation, R. W. Ketchledge, T. R. Finch. *AIEE Transactions*, vol. 72, pt. I, Sept. 1953, pp. 414–30.

3. The Development of Electron Tubes for a Coaxial System, G. T. Ford, E. J. Walsh. *Bell System Technical Journal*, New York, N. Y., vol. 30, Oct. 1951, pt. II, pp. 1103–28.

4. Coupling Networks, H. W. Bode. United States Patent No. *2337965*, Dec. 28, 1943.

5. Variable Equalizer, H. W. Bode. *Bell System Technical Journal*, New York, N. Y., vol. 17, April 1938.

6. The Potential Analogue Method of Network Synthesis, Sidney Darlington. *Bell System Technical Journal*, New York, N. Y., vol. 30, April 1951, pp. 315–65.

7. Network Synthesis Using Tchebycheff Polynomial Series, Sidney Darlington. *Bell System Technical Journal*, New York, N. Y., vol. 31, July 1952, pp. 613–65.

8. Network Analysis and Feedback Amplifier Design (book), Hendrik W. Bode. D. Van Nostrand Company, Inc., New York, N. Y., 1945

◆

No Discussion

Co-ordination of M1 and N1 Telephone Carrier Systems

E. P. SMITH
ASSOCIATE MEMBER AIEE

L. P. CORNELL, JR.
ASSOCIATE MEMBER AIEE

M. G. JEROME
ASSOCIATE MEMBER AIEE

Synopsis: The Bell System's new *N1* carrier system is a low-energy cable carrier system and is subject to interference from existing carriers with overlapping frequencies and higher energy levels. This paper is concerned with the development of a suppression device to prevent energy from the *M1* open-wire carrier system disturbing the channels of the *N1* system. The problem is complicated by the requirement that other types of open-wire carrier systems, notably the *01* system, must work through the suppression unit. Field tests of the suppression unit developed, consisting of a low-pass filter with mid-point grounded, indicate satisfactory transverse and longitudinal loss to the unwanted *M1* frequencies without degrading the performance of the line at frequencies below the *M1* band.

NEWLY developed methods of telephone transmission are not always completely different from existing techniques. This means that some engineering effort must be applied to co-ordinating the new methods with the old. In the case of the Bell System's new *N1* cable carrier system the co-ordination problems are considerable. The *N1* system frequency band overlaps those of the existing *K* cable carrier system, and the *J*, *M*, and *O* open-wire carrier systems. In addition to the frequency overlap, the systems operate at different levels, the *N1* having the lowest levels. Fig. 1 shows the frequency assignments of the various systems and indicates their relative transmitting levels.

Each of the conflicts among these systems presents its own peculiar problem. On some of the conflicts considerable work has been done, and much information is available. This paper treats only the problem of interference in *N1* systems caused by *M1* systems and the interrelated problem of operating *01* carrier on lines used for *M1* systems.

The M1 and N1 Conflict

The *M1* and *N1* systems are similar in that they are both double-side-band carrier-transmitted systems using different bands of frequencies for the two directions of transmission. The *N1* system frogs or interchanges the two frequency bands at each repeater. That is, if the high group of frequencies, as illustrated by Fig. 1, are received by the *N1* repeater, they are modulated by a group modulator with a frequency of 304 kc and transmitted to the next repeater as the low group. The high and low groups are kept separate at a repeater or terminal by a high low-pass filter which acts as a roofing filter for the low group. There is a pass-band filter at the input to the high-group repeater designed to suppress radio-frequency interference. However, the band-pass filter has little loss in the *M1* frequency range, being most effective above 500 kc. Under this arrangement, at the input to a high-group-to-low-group repeater, not only is there direct interference from the *M1* low-group frequencies, but in addition the *M1* high-group frequencies may be modulated by the 304 kc *N1* modulator and translated to frequencies within the *N1* band. Table I shows the *M1* carrier frequencies and the type of interference they produce in the *N1* channels.

Ordinarily at any one location there will be only one or two *M1* systems, but there may be 50 or more *N1* systems. Since each *M* channel can interfere with a channel in each *N* system, the interference must be controlled or there will be serious interference in many telephone facilities. The interference may be controlled by suppression, or the *M1* systems may be replaced with a new noninterfering type of system at appreciable economic penalty.

Suppression Required

As mentioned earlier, the *N1* system is a cable system and the *M1* an open-wire system. Where *M1* and *N1* high-group receive equipments were in the same office, experience indicated that even by maintaining good physical separation between the equipments, by using shielded wire for the carrier leads, and by providing separate cables for the *M1* and *N1* systems, there would still be excessive interference in the *N1* channels. The problem could be avoided by arranging the *N1* systems so that only low-group receive repeaters were located in offices containing *M1* equipments. However, the economics of additional *N1* repeaters, the problem of co-ordinating the direction of transmission of all the *N1* systems in an area, and the desirability of locating *N1* equipment in existing telephone buildings normally precludes the shifting of an *N1* repeater just to avoid conflict with a few *M1* channels. Where *N1* system layouts place *N1* high-group receive repeaters or terminals in the same location as existing *M1* equipment, approximately 125 decibels of loss is required between the *M1* carriers and the input to the *N1* equipment. The 125-decibel loss is based on the requirement that, to prevent excessive interference in an *N* channel, the level of an interfering tone should be 45 decibels below the level of the *N* channel carrier. Since the level of the *N* channel carrier at the input to a high-group repeater or terminal may be as low as 50 to 60 decibels below one milliwatt while the output level of the *M1* carrier is 25 decibels above one milliwatt, the required loss is 45+55+25, or about 125 decibels.

Paper 53-225, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 18, 1953; made available for printing May 1, 1953.

E. P. Smith, L. P. Cornell, Jr., and M. G. Jerome are with the Pacific Telephone and Telegraph Company, San Francisco, Calif.

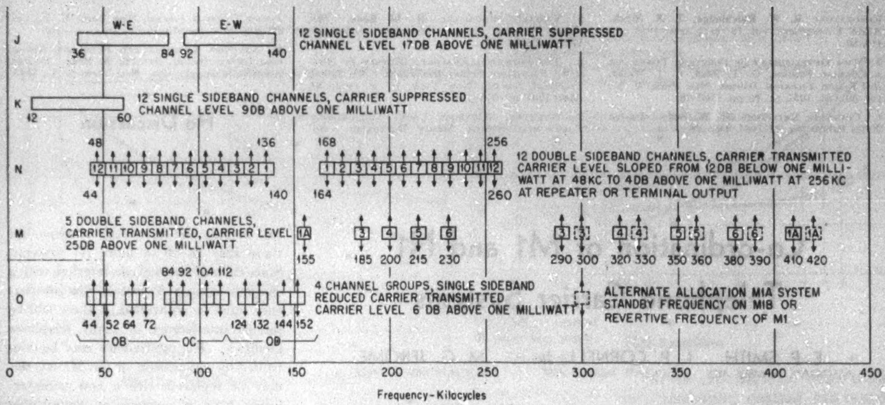

J
W-E
E-W
36
84 92
140
12 SINGLE SIDEBAND CHANNELS, CARRIER SUPPRESSED
CHANNEL LEVEL 17DB ABOVE ONE MILLIWATT
K
12
60
12 SINGLE SIDEBAND CHANNELS, CARRIER SUPPRESSED
CHANNEL LEVEL 9DB ABOVE ONE MILLIWATT
N
48
136
12 11 10 9 8 7 6 5 4 3 2 1
44
140
168
256
1 2 3 4 5 6 7 8 9 10 11 12
164
260
12 DOUBLE SIDEBAND CHANNELS, CARRIER TRANSMITTED
CARRIER LEVEL SLOPED FROM 12DB BELOW ONE MILLIWATT AT 48KC TO 4DB ABOVE ONE MILLIWATT AT 256KC
AT REPEATER OR TERMINAL OUTPUT
M
5 DOUBLE SIDEBAND CHANNELS,
CARRIER TRANSMITTED, CARRIER LEVEL
25DB ABOVE ONE MILLIWATT
1A 3 4 5 6
155 185 200 215 230
3 3 4 4 5 5 6 6 1A 1A
290 300 320 330 350 360 380 390 410 420
O
84 92 104 112
44 52 64 72
124 132 144 152
OB OC OD
4 CHANNEL GROUPS, SINGLE SIDEBAND
REDUCED CARRIER TRANSMITTED
CARRIER LEVEL 6 DB ABOVE ONE MILLIWATT
ALTERNATE ALLOCATION M1A SYSTEM
STANDBY FREQUENCY ON M1B OR
REVERTIVE FREQUENCY OF M1
0
50
100
150
200
250
300
350
400
450
Frequency-Kilocycles

Fig. 1. Frequency allocations

Nature and Estimated Loss of M to N Coupling Path

At these frequencies most of the coupling is on a longitudinal basis. The longitudinal component of the *M* energy couples through the telephone office wiring and onto the non-*N* pairs in the *N* cable where it crosstalks to the *N* pairs. On the *N* pairs the longitudinal *M* energy is converted to metallic *N* interference by the unbalances of the pairs. The longitudinal loss of the telephone office wiring is between 10 and 30 decibels. This, with a typical 10-decibel crosstalk coupling loss in the cable, and a longitudinal-to-metallic conversion loss of approximately 30 decibels, gives a total loss in the *M*-to-*N* coupling path of about 50 to 70 decibels. The loss in the coupling path is far short of the desired 125 decibels, as discussed. Additional loss may be obtained by moving the *M* terminals from the office to pole-mounted equipment cabinets on the open-wire line. With the *M* equipment pole-mounted, the loss in the coupling path is increased by the longitudinal loss of the open-wire entrance cable and the crosstalk coupling loss in the open wire. The longitudinal loss of cable at *M* frequencies is about 20 decibels per mile but, unfortunately, most entrance cables are relatively short, so that after moving the *M* terminals and allowing for 10 to 20 decibels crosstalk loss in the open wire there is usually some 20 to 40 decibels of loss still required. Fig. 2 shows a typical plant layout of the type just discussed.

The M1 and 01 Conflict

In past years, *M* carrier suppression devices have been developed for use on open-wire lines where it is desired to operate the same *M* channel at two different locations along the line. The two principal devices used by the Bell System are the type-*A* and type-*C* suppression units. Both type-*A* and type-*C* suppression units are half-section filters. Type *C* uses a different coil from that of type *A*, and has about 5 decibels more loss to *M1* frequencies. When inserted in an open-wire line both type-*A* and type-*C* suppression present an impedance mismatch to the open wire at frequencies above 30 kc. The resultant reflection coefficient is about 35 per cent at 70 kc to 70 per cent at 120 kc. Where it is desired to operate *01* carrier on a line equipped with type-*A* or type-*C* suppression, the high reflection contributes substantial reflected near-end crosstalk which will make some of the *01* carrier channels useless for commercial service. In addition, type-*A* suppression introduces a transmission impairment to *01* carrier frequencies which will reduce the allowable length of the *01* repeater section. Type-*C* suppression, which was designed for voice frequency lines, has a greater loss at carrier frequencies above 10 kc than does type-*A* suppression, with the result that neither *C* carrier nor *01* carrier can be used on a line equipped with type-*C* suppression units.

Table I. Interference from MI Carrier Systems into NI Carrier Systems

M1 Channel		Modulation Product, ±(304−M)*	N1 System		Interference in N1 System
Channel	Frequency, Kc(M)		Channel†	Channel Carrier, Kc	
1A	155	149			
	410	106	5L	104	2-kc tone§
	420	116	3L, 4L	120, 112	‡
3	185		3H	184	1-kc tone§
	290	14			
	300	4			
4	200		5H	200	intelligible crosstalk
	320	16			
	330	26			
5	215		7H	216	1-kc tone§
	350	46	12L	48	2-kc tone§
	360	56	11L	56	intelligible crosstalk
6	230		9H	232	2-kc tone§
	380	76	8L, 9L	72, 80	‡
	390	86	7L	88	2-kc tone§

* Frequency, kilocycles, resulting from *M* channel passing through *N* repeater.

† *L* = low group, *H* = high group.

‡ Possible interference to *N* signaling if frequency drifts.

§ Also unintelligible crosstalk.

The F1100 Filter

The problem resolves itself to developing a device that, when placed in an open-wire carrier line, will suppress, both longitudinally and metallically, *M* carrier energy by 30 decibels or more, and at the same time will not increase the transmission loss or the crosstalk of the line at fre-

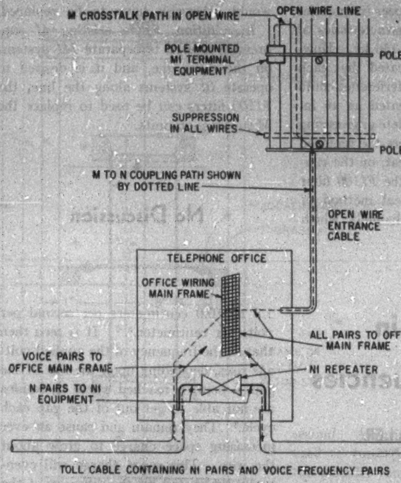

M CROSSTALK PATH IN OPEN WIRE
OPEN WIRE LINE
POLE MOUNTED M1 TERMINAL EQUIPMENT
POLE 2
SUPPRESSION IN ALL WIRES
POLE 1
M TO N COUPLING PATH SHOWN BY DOTTED LINE
OPEN WIRE ENTRANCE CABLE
TELEPHONE OFFICE
OFFICE WIRING MAIN FRAME
ALL PAIRS TO OFFICE MAIN FRAME
VOICE PAIRS TO OFFICE MAIN FRAME
N1 REPEATER
N PAIRS TO N1 EQUIPMENT
TOLL CABLE CONTAINING N1 PAIRS AND VOICE FREQUENCY PAIRS

Fig. 2. Typical M to N interference path

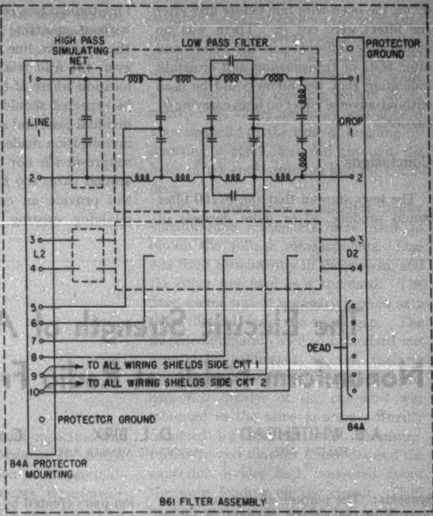

HIGH PASS SIMULATING NET
LOW PASS FILTER
PROTECTOR GROUND
LINE 1
DROP 1
1
2
3
4
L2
D2
5
6
7
8
9
10
TO ALL WIRING SHIELDS SIDE CKT 1
TO ALL WIRING SHIELDS SIDE CKT 2
DEAD
PROTECTOR GROUND
84A
84A PROTECTOR MOUNTING
B61 FILTER ASSEMBLY

Fig. 3. F1100 M1 suppression filter, 2 filters per B61 filter assembly box

quencies below the *M* range. In addition the installed cost of the device must be appreciably less than the cost of replacing the *M1* systems or adding an *N1* repeater point.

It was felt that a low-pass filter with the mid-points of certain shunt elements grounded would meet these requirements. The problem was discussed with representatives of an outside manufacturer who designed and built the *F1100* filter to the required specifications. The filter has a cutoff frequency of 175 kilocycles, and a longitudinal loss (on the bench with a zero-ohm ground) in the low-pass section of 50 or more decibels to frequencies above 180 kc. The transmission loss of the low-pass section was less than 0.3 decibel through the *01* carrier frequency range with a reflection coefficient when inserted in a 600-ohm line of less than 5 per cent in the passband. Two *F1100* filters are mounted with their associated protector strips in one type-*B61* filter assembly box for crossarm mounting. Fig. 3 shows a schematic of the filter and its wiring in the filter box.

The field trials involved only the suppression of *M* energy. Therefore, the *F1100* filters were ordered with the high-pass section replaced by a high-pass-section simulating network. This saved approximately $50 per filter. It is felt that when it is desired to operate *01* systems on the same pairs as the *M1* systems, the *M* coupling units may be replaced by the complete *F1100* filter, and the high-pass side of the filter used to couple the *M*'s to the line.

Results of Field Trials

Field trials were conducted at two locations, and longitudinal losses in excess of 30 decibels were obtained, see Table II. The arrangements necessary to obtain this loss are summarized:

1. A good local ground is required. The ground used for field trials had a d-c resistance of less than 2 ohms.
2. The shunt elements of the filter (Fig. 3) are terminated on separate lugs. This arrangement is provided so that separate ground leads can be used for each shunt element where maximum suppression is desired. Physically separated grounds for the shunt elements will reduce the amount of *M* energy that by-passes the series elements of the filter via the ground path. To obtain the results shown in Table II, terminals 5 and 7 were connected to a driven ground rod while terminal 9 was connected to the sheath of the entrance cable.
3. For the tests, all bridling and ground wires from the hot or line side of the filter were kept physically separated from the quiet or drop side leads. For greater suppression, it might be helpful to shield all leads to the filter.
4. The disturbing *M1* equipment was located far enough from the telephone office so that the effect of the run-around paths, near-by power lines, telephone exchange plant, fire alarm circuits, etc., was minimized. Where the possibility of the *M1* energy's by-passing the suppression unit via some adjacent wire structure is remote, the ideal location for the filter is on the terminal pole of the open-wire line between the open-wire line and the open-wire entrance cable. However, where the terminal pole supports one or more exchange cables as well as the toll entrance cable, it is not very satisfactory for a suppression point. Field trials indicated that where exchange cables were on the toll open-wire pole line, the best location for the suppression unit was at the end of the exchange cable. When the filter is placed in the open wire at the end of a cable, the cable sheath should be used for the shunt element ground with additional grounds provided by driven ground rods as required. The d-c resistance of the cable sheath and parallel ground rod to ground should not exceed 10 ohms.
5. The voice-frequency leads from the drop side of the *M1* equipment are suppressed with capacitors from each wire to ground. It was found that the *M1* equipment should be located one pole further from the office than the filters, and that the voice pairs to the *M1* should be shielded from the open wire, as well as suppressed.

Table II. Typical Suppression Given by F1100 Filter

Frequency, Kc	Suppression, Decibels
185	37
200	32
215	30
230	41
280	44
340	21
400	31

In addition to the suppression measurements, transmission and crosstalk tests were made in 2-kc steps from 40 kc to 156 kc on the open-wire line before and after the filters were cut in. There was no appreciable difference in the before and after tests of transmission and crosstalk indicating that the filters could be used without adverse effect on lines carrying *01* carrier.

Conclusions

The tests showed that the *F1100* filter would produce 30 or more decibels of longitudinal loss to *M* carrier frequencies without upsetting the characteristics of the open-wire line below 156 kc. However, the field trials indicated that each location where *M*-to-*N* interference could be expected should be treated as an individual case, and a complete engineering investigation made. Where the required suppression is not too great, or the run-around paths too many, the *F1100* filter will provide an economical method of retaining existing *M1* channels which would otherwise have to be replaced.

In addition, where existing *M* suppression is used to separate *M1* systems on the same line, and it is desired to operate *01* systems along the line, the *F1100* filters can be used to replace the *M* suppression units.

◆

No Discussion

The Electric Strength of Air in Nonuniform Fields at Radio Frequencies

J. B. WHITEHEAD
FELLOW AIEE

D. L. BIRX
ASSOCIATE MEMBER AIEE

C. F. MILLER
MEMBER AIEE

Synopsis: The purpose of this study is to investigate the electric strength of air in the frequency range of 2.0 to 16.0 megacycles per second, using a central wire and cylinder as electrodes. Work has previously been done with a similar electrode arrangement up to 1.5 megacycles per second.[1] A reduction in breakdown voltage up to 6.0 per cent was noted in this instance. The present investigation shows that above 2.0 megacycles per second, the breakdown voltages are an average of 32.0 per cent under the corona starting voltages, at 60 cycles per second.

Theory of Breakdown

THE ORIGINAL theory of breakdown as proposed by Townsend[2] is still basically accepted today, but with some modifications.[3] This theory is based on the premise that a free electron within the gap will be accelerated by the field, and if the field is strong enough this electron may acquire sufficient energy to ionize a molecule upon collision. The newly formed electron travels on with the original and in this way an avalanche of electrons is formed.

Behind the avalanche is left a trail of positive ions. In the case of d-c fields or at low power frequencies, when the field due to the concentration of these ions reaches a certain fraction of the applied field, a positive space-charge streamer will be formed, which is capable of self propagation towards the cathode.[4] The conditions in such an avalanche in an undistorted field may be expressed in terms of the distance, x, of the electron from its starting point and Townsend's first coefficient α which is the number of ion pairs created by one electron per unit length of its path in the field direction. For one electron starting, n electrons appear x centimeters nearer the anode, where, for a uniform field

$$n=e^{\alpha x}$$

and for a nonuniform field

$$n=e^{\int \alpha dx}$$

With cylindrical electrodes, it would seem reasonable that the electron must attain its critical velocity when moving toward the wire, because the intensity of the field increases in that direction. Thus it has been observed that corona starts first when the wire is positive.[5] At the critical voltage, the wire breaks out into a glow known as visual corona.

If the voltage is raised above this point, the boundary of the glow extends out in space towards the outer cylinder, and eventually a voltage is reached when a highly ionized column of air extends completely across the gap, and a spark breakdown occurs. The ionic charge formed around a corona wire during any half-wave is all of the same sign as the potential of the wire and may be referred to as a space charge. The boundary of this space charge varies with the frequency, size of wire, and rise of voltage above the corona voltage.[5]

At power frequencies, the space charge formed on any half-wave returns to the wire on the next succeeding half-wave. The velocity of the ions varies with the strength of the field. Values of mobility are reported for positive ions ranging from 1.5 to 10.0 centimeters per second per volts per centimeter.[5–7] It is seen then that as the frequency of the applied voltage rises, depending upon the size of the gap, a point is reached where these ions are not able to get out of the gap each cycle.[8] They remain and cause an ever increasing space charge to grow about the wire. This space charge will eventually be limited by the diffusion of the ions towards both electrodes, where recombination can readily take place. A reduction in the total voltage across the electrodes at which breakdown occurs is therefore possible, not only because the concentration of positive ions that has to be produced by the electron avalanche initiating breakdown will be less, but also because the initial space charge is sufficient to distort the electric field within the gap and so produce enhanced values of α.

Experimental work has shown that at still higher frequencies, another reduction in the breakdown voltage does occur,[9] and this is attributed to the fact that the electrons find it impossible to get out of the gap at this stage.

The peculiarities of radio-frequency discharges support this theory quite well. It is observed that over the frequency range herein investigated, corona as it is known at 60 cycles per second does not exist. No visible effect is noticed until the voltage is reached which causes complete breakdown of the gap between the wire and cylinder. The discharges occur

---

Paper 53-262, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted October 29, 1952; made available for printing April 30, 1953.

J. B. Whitehead and C. F. Miller are with The Johns Hopkins University, Baltimore, Md., and D. L. Birx is with The Franklin Institute, Philadelphia, Pa.

The authors gratefully acknowledge their indebtedness to Dr. R. K. Witt and Dr. J. J. Chapman of The Johns Hopkins University Institute for Cooperative Research for the use of their dielectric test equipment in making the measurements.

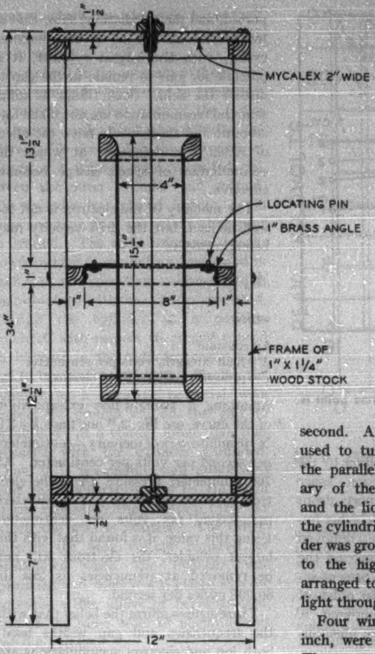


Fig. 1. Test setup showing mount for wire and cylinder

at one or several points along the wire, and take the form of brushes extending radially out from the wire. The breakdown is complete, extending all the way across the gap for the wire sizes and cylinders used. This effect has been observed before.[10]

If breakdown occurs initially on the part of the wave when the wire is positive, the accumulated positive space charge will have the effect of extending the inner conductor out toward the surrounding cylinder, causing most of the voltage to be applied between the cylinder and some point intermediate between the cylinder and the wire. When a discharge occurs, its tendency then is to completely bridge the gap.

Experimental Procedure

A Navy type-*TDH-4* transmitter with power capability of 3 to 5 kilowatts over the frequency range of 2 to 18 megacycles per second was used to provide the breakdown voltages. Four radio-frequency transformers provided voltage step-up, and these were tuned at frequencies of 2.0, 4.8, 9.81, and 16 megacycles per second. A variable liquid capacitor was used to tune the secondaries.[11] Across the parallel combination of the secondary of the radio-frequency transformer and the liquid capacitor was connected the cylindrical electrode gap. The cylinder was grounded, and the wire connected to the high voltage. A sunlamp was arranged to throw a beam of ultraviolet light through the gap.

Four wires, nos. 23, 16, 12, and 1/8 inch, were used for the investigation. These were suspended in turn within the cylinders, nos. 1, 2, 3, and 4, of radii 1.15, 2.44, 3.30, and 4.9 centimeters respectively. The length of the cylinders was in every case at least four times the diameter, to assure uniform field at the center. Grading flares were provided at the ends. The electrodes were mounted in a sturdy wooden frame with glass bonded mica insulation at the ends to support the wire. The cylinders were designed to mount each on the same support whereon they were located by means of aligning pins. This arrangement permitted rapid change of electrodes while at the same time assuring perfect alignment. The entire assembly was aligned very carefully at the time of building by inserting a rigid steel rod through the holes in the end insulators and the cylinders.

Two combination capacitor-divider vacuum-tube voltmeters were used to obtain the voltage measurements. One was fixed permanently in the system, and the other was used as a standard. The fixed meter was of necessity located at a finite distance from the spark gap. The meter combination used as a standard was accurately calibrated at the factory. During calibration of the fixed meter, the capacitor divider of the standard was mounted in the same position normally occupied by the electrodes. Measurements showed that the electrodes and the capacitor divider each presented about the same loading on the circuit. Calibration runs were made for each frequency used.

The high-Q resonant circuits used to generate the breakdown voltages provided a means of detecting the start of any discharge. The voltage was applied by slowly reducing the bias on the final stage of the transmitter, and breakdown was indicated by a sudden fall in the applied voltage, as the circuits were extremely sensitive to any loss.

The breakdown voltage was found to be especially affected by the surface condition of the wire. Before each measurement the wires were polished with jeweler's rouge, wiped with a rag dipped in alcohol and then washed with distilled water. The inside surfaces of the cylinders were polished, but dust or slight

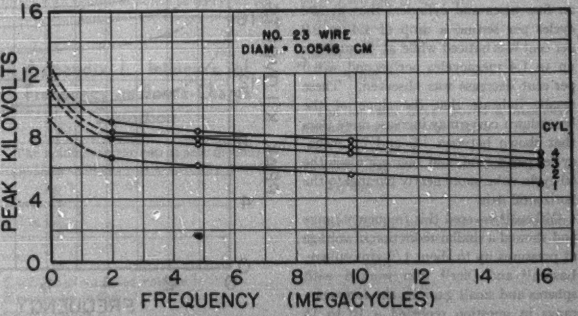


Fig. 2. Gap breakdown voltage using no. 23 wire as central electrode. 60-cycle point is voltage for visual corona

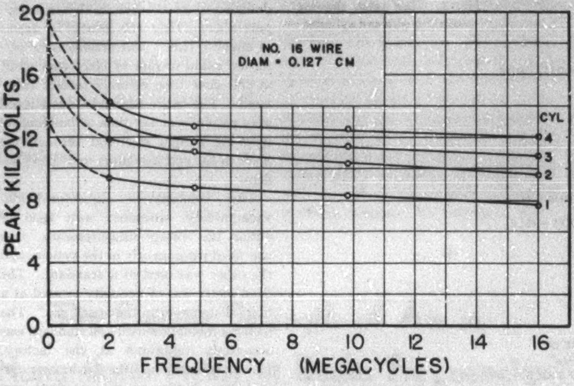


Fig. 3. Gap breakdown voltage using no. 16 wire as central electrode. 60-cycle point is voltage for visual corona

irregularities here were found to have little effect on the voltage.

Results

The results of the investigation are shown in Figs. 2 to 5. The breakdown voltage versus frequency curves are for the most part straight lines, all having a slight negative slope. This slope has a tendency to be greater at the lower frequencies. The experimental curves are connected with dotted lines to the 60-cycle voltage corresponding to the appearance of visual corona. The exact shape of the curve between 60 cycles and 2.0 megacycles per second is not known, but it probably does not differ much from the form shown.

Whitehead and Gorton[12] found that corona voltage at 2.0 kc is lower by 3 to 4 per cent than it is at 60 cycles per second. Misere[1] reported in experiments on fine wires and cylinders that at 500 cycles per second a drop of 1.4 to 2.0 per cent was noticed while at frequencies up to 1.4 megacycles per second, a 6.0 per cent decrease was observed. These results indicate that the slope of the breakdown curve may increase more than that shown between 1.0 and 2.0 megacycles per second, and then approach the 60-cycle value more nearly parallel to the horizontal axis.

Githens[13] covered this frequency range and showed a similar reduction in voltage at pressures up to about 1/2 atmosphere. Lassen[14] and Oler[15] who worked with spheres and small gaps in the frequency range in question reported a 10 to 15 per cent lowering of the radio frequency breakdown voltage from the 60-cycle value. This drop is accounted for by the accumulation of a positive-ion space charge in the gap. Pim[9] reports drops of approximately 44 per cent at higher frequencies, where electrons are not removed from the gap and remain to increase the ionization efficiency. In the present investigation, where drops of the order of 30 to 35 per cent are encountered, it would appear that both of these phenomena are active.

In Appendix I, the frequencies at which the ions and the electrons are left in the gap are computed. Two ion mobilities are used to cover the range of those reported, and the calculations show that at frequencies as low as several hundred cycles, with the largest cylinder, it is possible for ions to remain in the gap to distort the field. Such things as diffusion and recombination are not taken into account and these would have the effect of raising the frequency at which the accumulation of space charge becomes effective.

The mobility of the electron is not too well defined, but the drift velocity may be expressed as

$$v = k\frac{X}{P}$$

where

k = constant
X = field strength, volts per centimeter
P = pressure, millimeters of mercury

Assuming a straight-line extrapolation of the curve, see Fig. 2,[16] one finds $k = 3.9 \times 10^5$ millimeters of mercury = centimeters per second per volts per centimeter. At 760 millimeters pressure, then, the electron mobility would be 513 centimeters per second per volts per centimeter. Using this value, it is found that with the largest cylinder, the electrons will not be removed at frequencies as low as 60,000 cycles per second.

These figures affirm the belief that with the frequencies and gap widths used, both ion and electron accumulations are contributing to the lowering of the voltage. The downward slope of the curves out to 16 megacycles per second would indicate that the point has not been

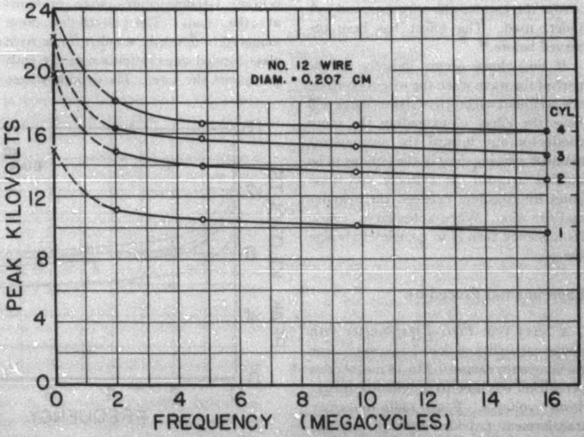


Fig. 4. Gap breakdown voltage using no. 12 wire as central electrode. 60-cycle point is voltage for visual corona

reached where the electrons have their maximum effect and the breakdown voltage levels off.

As a check on the accuracy of the theory of breakdown it was desired to determine, if possible, on what part of the radio-frequency cycle discharge actually occurred. To accomplish this, a source of steady d-c voltage was connected between the outer cylinder and ground. The cylinder was placed first at +430 volts and then at −430 volts with respect to ground. The radio-frequency breakdown voltage to ground was measured in each case, and this value was compared to that obtained with the cylinder grounded. Raising the cylinder to a positive potential with respect to ground caused an increase in the radio-frequency voltage to break down the gap. Placing the cylinder at a negative potential with respect to ground caused the gap to break down at a voltage lower than that required with the cylinder grounded. Thus, it is reasonably well established that the discharge occurs on that half of the radio-frequency voltage cycle for which the wire is positive. The difference in the peak voltage required for breakdown when the cylinder was grounded and when it was placed at a d-c potential either above or below ground, was of the same order of magnitude as the d-c voltage.

The absolute accuracy of the voltage measurements in this investigation is limited by the accuracy of the radio-frequency voltmeter used as a standard. This meter was carefully calibrated at the factory, and corrections, where necessary, were supplied by the manufacturer. Each point on the breakdown curves was determined by the average of at least five readings. Except for the smallest wire, the average deviation of these readings was less than ±2.0 per cent. Measurements on different days showed that the results are repeatable within a limit of ±2.0 per cent.

Conclusions

1. Accurate measurements of the breakdown potential of atmospheric air at radio frequencies between central wire and cylinder electrodes have been made. A frequency range of 2.0 to 16.0 megacycles per second was investigated, disclosing that the apparent strength of air at 2.0 megacycles per second is approximately 30 per cent below the 60-cycle value, and that the breakdown voltages decline at an average linear rate of 120 volts per megacycle as the frequency is raised above this point.

Table I. Critical Frequencies

Wire Size	Cylinder Radius R, Centimeters	Critical Frequency, Kc		
		Ions		Electrons
		$K_i = 1.5$	$K_i = 8.0$	$K_e = 513$
1/8 inch	1.15	6.58	35.1	2250.0
No. 16	1.15	3.23	17.2	1100.0
1/8 inch	4.90	0.35	1.9	121.0
No. 16	4.90	0.18	0.9	60.6

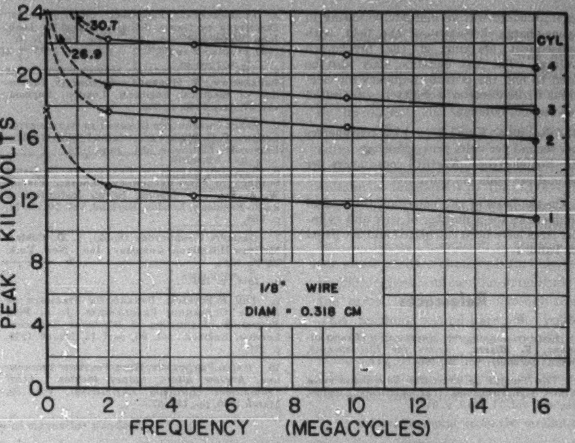


Fig. 5. Gap breakdown voltage using 1/8-inch wire as central electrode. 60-cycle point is voltage for visual corona

2. A modified Townsend theory of breakdown can be used to explain the results obtained. It has been demonstrated that due to the finite mobilities of the electrons and ions, these particles will not be cleared from the gap, but will remain to cause field distortion which results in lower breakdown voltages and discharges of a different nature from those obtained at 60 cycles per second.

3. The discharge was found to occur on that part of the cycle when the wire is positive with respect to the cylinder. This agrees with results obtained at 60 cycles per second, and supports the theory presented.

Appendix I. Influence of Frequency on Space-Charge Distortion

The gradient in the gap at any distance x from the axis of the wire and at any time t is

$$X_x = \frac{E}{x \log_e \frac{R}{r}}$$

where

$E = E_0 \sin \omega t$ = applied voltage
R = cylinder radius
r = wire radius

The velocity v of an ion at point x is then given by

$$v = \frac{k\, E_0 \sin \omega t}{x \log_e \frac{R}{r}}$$

where

k = mobility constant of ion

The distance dx traversed in time dt is then

$$dx = \frac{k\, E_0 \sin \omega t}{x \log_e \frac{R}{r}}\, dt$$

Integrating

$$\frac{x^2}{2} = -\frac{k\, E_0 \cos \omega t}{\omega \log_e \frac{R}{r}} + C$$

Now assume ions to be formed at the surface of the wire and have zero initial velocity at the point in the voltage cycle when $E = 0$ and $\omega t = 0$, $v = 0$.

Then

$$C = \frac{r^2}{2} + \frac{kE_0}{\omega \log_e \frac{R}{r}}$$

The ion will have moved a maximum distance x when $\omega t = \pi$. Here

$$x^2 = \frac{4kE_0}{2\pi f \log_e \frac{R}{r}} + r^2$$

Now placing $x = R$, the cylinder radius, and solving for the frequency

$$f_c = \frac{2kE_0}{\pi (R^2 - r^2) \log_e \frac{R}{r}}$$

This is the critical frequency, above which not all of the ions will be removed from the gap. Charge will accumulate, resulting in space-charge distortion of the field, with subsequent lowering of the breakdown voltage. The last equation may also be used to find the critical frequency for electrons if the electron mobility is substituted for the ion mobility.

1. For ions, $K_i = 1.5$ to 10.0 centimeters per second per volts per centimeter.
2. For electrons, $K_e = 513$ centimeters per second per volts per centimter

On the basis of this equation, the critical frequencies for a number of wire and cylinder sizes have been calculated and are listed in Table I.

References

1. Gasentladung bei Neider-und Hochfrequenz, **F. Misere.** *Archiv für Electrotechnik*, Berlin, Germany, vol. 28, 1934, p. 411.

2. The Theory of the Glow Discharge from Wires, **J. S. Townsend.** *The Electrician*, London, England, vol. 71, 1913, p. 348.

3. Fundamental Processes of Electrical Discharge in Gases (book), **Leonard B. Loeb.** John Wiley and Sons, Inc., New York, N. Y., 1939, chapters 8 and 9.

4. The Positive Streamer Mechanism of Spark Breakdown, **W. Hopwood.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 97, July 1950, p. 253.

5. Space Charge and Current in Alternating Corona, **C. H. Willis.** Dissertation, Johns Hopkins University, Baltimore, Md., June 1926.

6. The Initiations of Breakdown in Gases Subject to High-Frequency Electric Fields, **W. A. Prowse.** *Journal*, British Institution of Radio Engineers, London, England, vol. 10, no. 11, p. 333.

7. Gaseous Conductors (book), **J. D. Cobine.** McGraw-Hill Book Company, Inc., New York, N. Y., 1941.

8. *Ibid.*, p. 185.

9. The Electrical Breakdown Strength of Air at Ultrahigh Frequencies, **J. A. Pim.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 96, part II, March 1949, p. 117.

10. Radio-Frequency High-Voltage Phenomena, **Andrew Alford, Sidney Pickles.** *AIEE Transactions* (*Electrical Engineering*), vol. 59, March 1940, pp. 129–37.

11. Testing Dielectric Properties at High Frequency, **R. K. Witt, J. J. Chapman.** *Tele-Tech*, New York, N. Y., Dec. 1950.

12. The Electric Strength of Air—V, **J. B. Whitehead, W. S. Gorton.** *AIEE Transactions*, vol. XXXIII, 1914, pp. 951–72.

13. The Effect of Frequency on the Corona, **W. S. Gorton.** Dissertation, Johns Hopkins University, Baltimore, Md., June 1914.

14. The Influence of Discharge Chamber Structure upon the Initiating Mechanism of the High-Frequency Discharge, **S. Githens.** *Physical Review*, New York, N. Y., vol. 57, 1940, p. 822.

15. Frequenz Abhängigkeit der Funkenspannung in Luft, **H. Lassen.** *Archiv für Elektrotechnik*, Berlin, Germany, vol. 25, 1931, p. 332.

16. The Use of Sphere Gaps at Radio Frequencies, **C. B. Oler.** Dissertation, Johns Hopkins University, Baltimore, Md., June 1950.

17. Electron and Negative Ion Mobilities in Oxygen, Air, Nitrous Oxide, and Ammonia, **H. A. Neilsen, N. W. Bradbury.** *Physical Review*, New York, N. Y., vol. 51, 1937, p. 69.

◆

No Discussion

Preliminary Development of a Magnettor Current Standard

E. P. FELCH
MEMBER AIEE

J. L. POTTER
MEMBER AIEE

Synopsis: In the wartime development of the air-borne magnetometer, a method of detecting extremely small changes in magnitudes of magnetic fields was developed. The principle involved was the use of a second-harmonic type of magnetic modulator now known as a magnettor. This instrument can detect changes in magnetic fields in the order of 10^{-6} oersted. A study was made at Rutgers University under the sponsorship of Bell Telephone Laboratories to determine the feasibility of obtaining a standard of current using the magnettor principle. In this case, the magnettor was placed in a stable permanent-magnet field of a few oersteds and the field nulled by means of a solenoid carrying the current to be stabilized. This current was supplied by an electronic self-balancing system, which maintained a null condition. If a field of 1 oersted is used, an ultimate accuracy of one part in 10^6, or 0.001 per cent, might be expected.

In the experimental tests, when the temperature was varied over a range from 20 to 50 degrees centigrade, an accuracy of better than 0.1 per cent was maintained, and if the magnettor was held at a constant temperature the accuracy was better than 0.01 per cent.

The device can be used to develop a standard of current and voltage. It can also be used as a power level reference standard or as a voltage reference source in regulated power supplies where high accuracy and stability are demanded. It is believed that with further development work the instrument can be made to approach the accuracy of the standard cell.

IN THE development of electric equipment and in electrical measurements, it is desirable to have standards of voltage and current for use as power reference standards, in communications systems, in calibration of indicating instruments, as standard reference voltages for regulated power supplies, and as potentiometer standardizing references.

The panel type of indicating instrument has an accuracy of ±2 per cent of full-scale value, the portable laboratory instrument has an accuracy of 0.25 to 1 per cent, and the best laboratory standard indicating instrument has an accuracy of 0.1 per cent of full-scale value. These laboratory standards are not considered portable as they require leveling and special handling.

The standard cell used with a suitable measuring potentiometer is the ultimate in accuracy in measuring electric voltage or current. The accuracy is about 0.001 per cent if extreme care is taken. The normal accuracy is better than 0.01 per cent.

The gas discharge tube is commonly used as a reference voltage source in regulated power supplies. These tubes will maintain a voltage within 1 or 2 per cent, and special tubes may be better than 0.25 per cent for a relatively short period of time.

Rutgers University entered into a contract with the Bell Telephone Laboratories to study the feasibility of using the second-harmonic type of magnetic modulator, known as a magnettor, in conjunction with a standard magnetic field to produce a current or voltage standard for reference purposes that would be a portable, rugged, and reliable device.

The theory of the operation of the magnettor is described in the literature[1-6] and will not be treated in this paper. The application of such techniques to a current standard was suggested several years ago.[7]

Electric Circuitry

Fig. 1 is a block diagram showing the circuit arrangement used during these experimental tests. The 1,000-cycle os-

---

Paper 53-291, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15-19, 1953. Manuscript submitted December 15, 1952; made available for printing May 11, 1953.

E. P. Felch is with the Bell Telephone Laboratories, Inc., Murray Hill, N. J., and J. L. Potter is with Rutgers University, New Brunswick, N. J.

Sylvan Fich, W. R. Carson, and R. C. Pegg made important contributions to this project at Rutgers University. Special thanks are due J. G. Ferguson, V. E. Legg, and F. G. Merrill of Bell Telephone Laboratories for their continued interest and the benefit of their consultation during the period of the work.

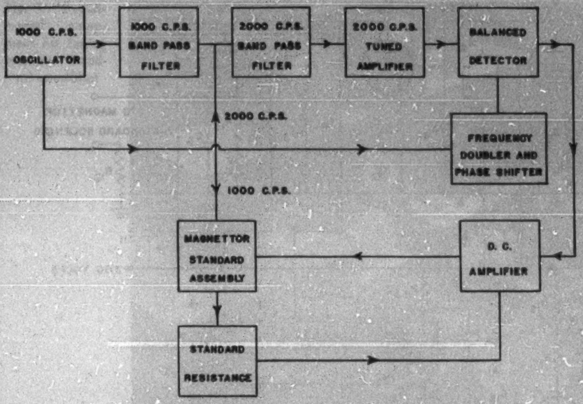

1000 C.P.S. OSCILLATOR
1000 C.P.S. BAND PASS FILTER
2000 C.P.S. BAND PASS FILTER
2000 C.P.S. TUNED AMPLIFIER
BALANCED DETECTOR
2000 C.P.S.
FREQUENCY DOUBLER AND PHASE SHIFTER
1000 C.P.S.
MAGNETTOR STANDARD ASSEMBLY
D. C. AMPLIFIER
STANDARD RESISTANCE

Fig. 1. Block diagram of magnettor standard

cillator has its wave form purified by a 1,000-cycle band pass filter which excites the magnettor standard assembly located inside a magnetic shield case. This cavity contains the permanent magnets and d-c solenoid which produce opposing fields on the magnettor core. If a d-c residual flux is produced in the magnettor core, a second harmonic is produced across the magnettor coil and is accepted by a 2,000-cycle filter. This signal is amplified by a 2,000-cycle tuned amplifier. The 2,000-cycle output from the amplifier will reverse in phase 180 degrees when the d-c flux in the magnettor core reverses direction.

To detect this phase change in voltage, the 1,000-cycle voltage from the oscillator is doubled in frequency in the frequency doubler and then, by a phase-shifter circuit, can be brought in phase or 180 degrees out of phase with the 2,000-cycle error voltage of the amplifier.[2,3,9] These two signals are combined in a balanced detector circuit and rectified so that a direct voltage output is produced that is positive for one phase of the 2,000-cycle error voltage and negative for the opposite phase. The output of the d-c amplifier, which is a cathode-follower in this case, feeds the direct current into the d-c solenoid which opposes and nulls the field of the permanent magnets. If a complete nulling of the two fields were produced, the 2,000-cycle error voltage would be zero. The current flowing in the standard resistance is checked by means of a conventional precision potentiometer-measuring circuit (not shown in diagram).

The circuit diagram of the amplifier and phase detector as used for these tests on the magnettor and cavity is shown in Fig. 2. The amount of development work done on the circuit was limited, and it could undoubtedly be simplified and improved. The voltage gain of the a-c amplifier was about 82 decibels, which was adequate for these tests.

The experimental laboratory setup for testing the magnettor standard is shown in Fig. 3 with a portion of the auxiliary test equipment. The magnet magnetizing equipment and wave analyzer are not shown in this setup. Preliminary work was done using a wave analyzer in place of the amplifier, and current was supplied to the solenoid by using a battery and variable resistance.

Experimental Work and Results

Considerable initial thought was given to possible structures that could be used to produce and maintain a standard magnetic field for the magnettor unit. It was decided to try relatively long rod magnets mounted in a retaining magnetic case or cavity. The long magnets were considered advantageous because they would not be demagnetized easily.[3]

A number of the experimental cavities which were constructed during the course of the work are shown in Fig. 4. The first of these cavities was constructed using Alnico magnets and is shown as the second from the top in Fig. 4. When this cavity was tested and the magnet field was nulled by a direct current from a battery, it was found that the smallest output voltage that could be obtained was about 1,200 microvolts output from the 2,000-cycle filter. When the magnettor was nulled in the earth's field, the residual second-harmonic output was about 10 microvolts. Hence it was realized that the shape of the flux field produced by the magnets and the cavity must have a major effect upon the residual second-harmonic voltage.

When the phase of the second-harmonic voltage is compared to the 1,000-cycle driving signal, the second-harmonic phase remains unchanged at high signal outputs, but as the null condition is approached the phase position changes from 0 degrees to 180 degrees without the magnitude of the error voltage going through zero. This residual second-harmonic voltage can then be considered as a quadrature component of second-harmonic voltage. The quadrature component of second-harmonic voltage may be due to:

1. a second harmonic from the oscillator;
2. stresses in the magnettor core;
3. the two magnetic fields (that from the magnet and that from the solenoid) having flux lines of dissimilar shape.

The reason why item 3 produces the quadrature component has not been definitely determined. When a solenoid longer than the core nulls the earth's field, the lines of flux of both fields are essentially parallel along the axis of the magnettor core, and do not produce an appreciable quadrature component in this case.

Since difficulty was experienced with the quadrature component in the initial cavity constructed, a study was made with an experimental setup in a large magnetic shield case. This model was made of plastic material with magnets held in place by tape on a plastic tube. This unit is shown as the bottom unit in Fig. 4. Numerous tests were conducted with the plastic model to determine the cause of the quadrature component, such as:

1. nulling one solenoid field against another;
2. changing the number of magnets;
3. changing the length of the magnets;
4. changing the strength of the magnets.

The nulled field of the two solenoids made no noticeable difference as long as they were concentric and symmetrical with respect to the magnettor core. When one solenoid was moved up or down with respect to the axis, the residual second-harmonic voltage increased.

The most significant point was the effect that the change in magnet length had on the residual voltage. This is shown for a typical case in Fig. 5. It is

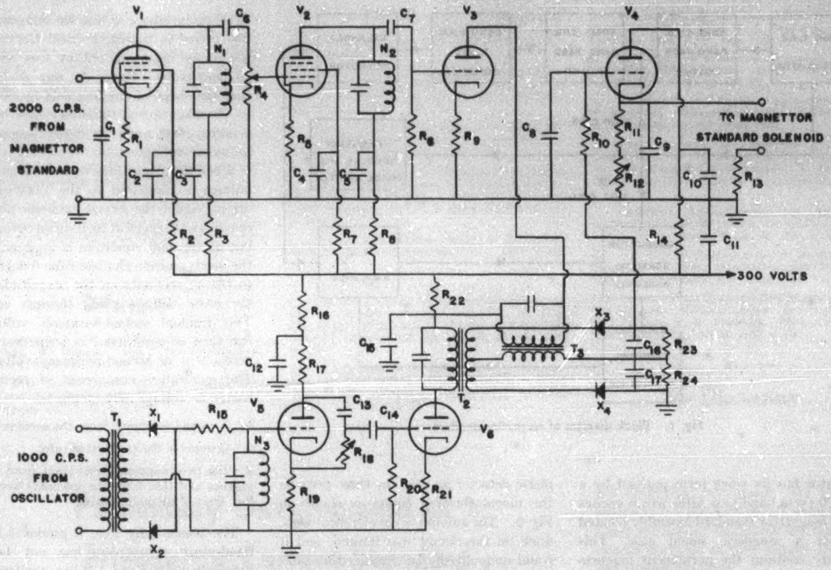


Fig. 2. 2,000-cycle-per-second amplifier and phase detector. The component list is:

Part	Description	Part	Description	Part	Description
C_1	20 micromicrofarads	R_6	250,000 ohms	N_1, N_2, N_3	2,000 cycle-per-second tuned circuit
C_2, C_3, C_4, C_5,		R_9, R_{21}	2,000 ohms		
C_{11}, C_{12}, C_{15}	0.25 microfarad	R_{10}, R_{20}	1 megohm	T_1	5,000 ohms: 5,000-ohm transformer center-tap secondary
C_6, C_7, C_{13}	0.005 microfarad	R_{11}	10,000 ohms		
C_8	10 microfarads	R_{12}	5,000 ohms variable	T_2	primary 0.1 henry tuned to 2,000 cycles per second; secondary 0.1 henry center tap
C_9, C_{16}, C_{17}	0.1 microfarad	R_{13}	25,000 ohms+1,000 standard resistance		
C_{10}	1 microfarad	R_{14}, R_{16}, R_{22}	10,000 ohms		
C_{14}	0.01 microfarad	R_{15}	50,000 ohms	T_3	same as T_2 except center tap may be omitted
R_1, R_5	200 ohms	R_{17}, R_{19}	5,000 ohms		
R_2, R_3	80,000 ohms	R_{18}	100,000 ohms variable	V_1, V_2	*6AK5*
R_8, T_2	20,000 ohms	R_{23}, R_{24}	100,000 ohms	V_3, V_6	1/2 *12AU7*
R_4	250,000 ohms variable	X_1, X_2, X_3, X_4	*1N34* rectifier	V_4, V_5	*6C4*

noted that there is an optimum length of magnet for this particular construction.

A rather extensive secondary effect was the change in residual voltage with the strength of the magnets. A typical curve is shown in Fig. 6. This was not expected but was observed in the previous tests. This change was attributed to the fact that the kernels of the magnetic poles shifted up and down along the rod magnets as the strength of the magnets was varied. Drill-rod magnets were used in all of these tests.

Factors affecting choice of magnet length are listed in Appendix I.

A second cavity was constructed and is shown in Fig. 4, second from the bottom. This cavity was constructed so that magnets of different lengths could be used. Tests were made with two magnets and four magnets with identical results. It was found that the residual second-harmonic voltage could be reduced to a satisfactory level using this assembly. Little change was required in magnet length in going from the plastic model to the cavity to keep the residual second-harmonic voltage at a low value. When a shift was made from the steel drill-rod magnets to Vicalloy magnets, a small change in magnet length had to be made in order to obtain a satisfactory nulling level. This again was attributed to shifting of pole kernels.

It might be stated at this point that two lengths of magnettor cores were tried. The lengths were 3/4 inch and $1^1/_2$ inches. No significant difference in results was observed except that the short core had considerably less output. Hence most tests were conducted with the $1^1/_2$-inch cores.

After eliminating the quadrature component of second-harmonic voltage to a satisfactory degree, attention was turned to stability of the field with respect to time and temperature; especially to determine any hysteresis effect as a function of temperature. Several tests were conducted with this cavity using drill-rod and Vicalloy magnets, which were supported on the same brass frame as the d-c solenoid and magnettor. The Vicalloy magnets were, of course, superior in stability, but the results showed a temperature coefficient that was negative. This cavity was then modified to the extent

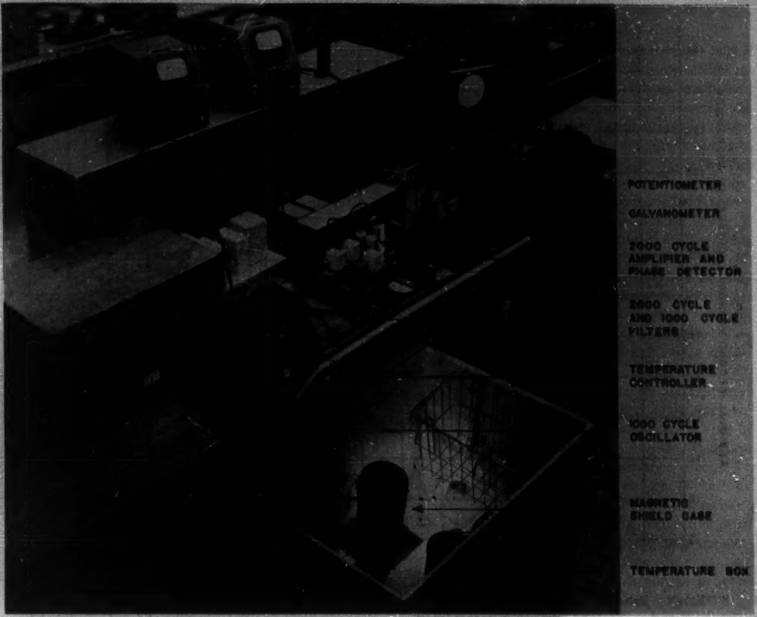


Fig. 3. Experimental equipment for testing magnettor standard assemblies

of mounting the magnets in two brass holders fastened to the cavity wall. This construction showed an improvement in temperature coefficient because the brass holders and the steel cavity expanded in the opposite directions. This resulted in a decrease in the negative temperature coefficient.

It was reasoned from the results of the previous tests that a larger diameter cavity would further decrease the temperature coefficient. A new cavity was designed and constructed with a larger diameter. It is shown in Fig. 4 as the top cavity. It was also decided to shorten the length of this cavity. It was found in testing the cavity that the residual second-harmonic voltage could not be reduced to a satisfactory value regardless of the length of the magnets used. This result was attributed to the shorter length of the cavity. However, several tests were made with the cavity and it was found to have a positive temperature coefficient.

The results of these tests led to the design and construction of another cavity, shown in Fig. 4 as the middle assembly. This cavity was longer than the preceding one and of a smaller diameter. Tests made with it indicated that it had a negative temperature coefficient that was lower than any obtained with previous cavities. The greater length also made possible low values of residual second-harmonic voltage. In all these cavities, little change in the length of the magnets was required to give satisfactory results.

Considerable difficulty was experienced in early tests with the current drifting in value with respect to time. This condition was found to be caused by residual magnetism in the cavity which was being demagnetized in the presence of the earth's field. When the final demagnetization was made inside a

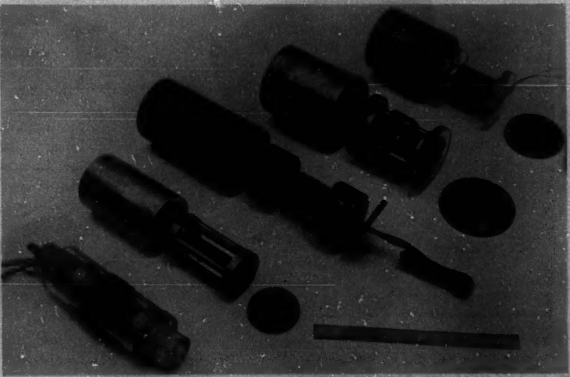


Fig. 4. Experimental magnettor cavities and assemblies

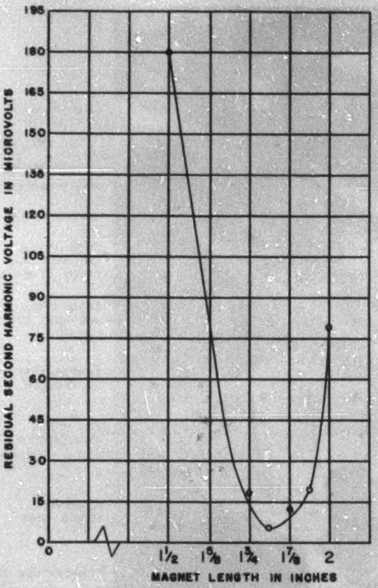
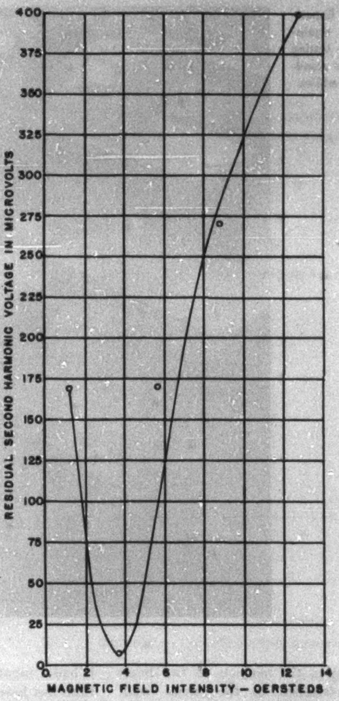


Fig. 5 (above). Residual second-harmonic output voltage from 2,000-cycle-per-second filter versus magnet length

Fig. 6 (right). Residual second-harmonic output voltage from 2,000-cycle-per-second filter versus magnetic field intensity of the magnets



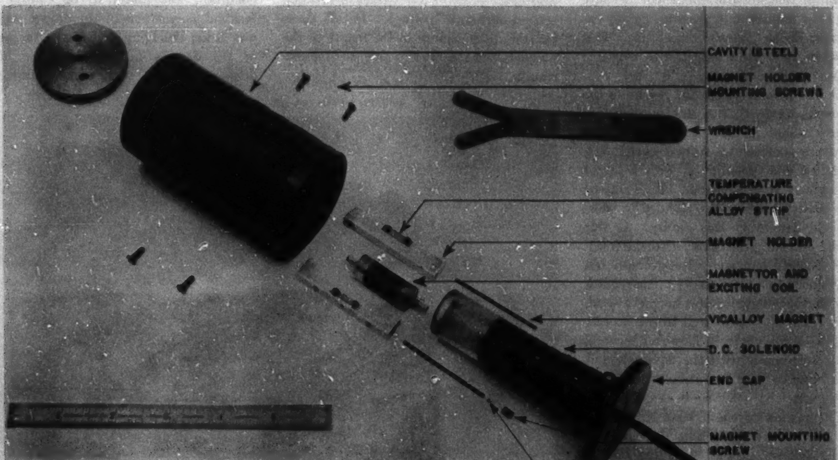

CAVITY (STEEL)
MAGNET HOLDER MOUNTING SCREWS
WRENCH
TEMPERATURE COMPENSATING ALLOY STRIP
MAGNET HOLDER
MAGNETTOR AND EXCITING COIL
VICALLOY MAGNET
D.C. SOLENOID
END CAP
MAGNET MOUNTING SCREW
MAGNET MOUNTING SPRING

Fig. 7. Cold-rolled steel magnettor cavity and assembly

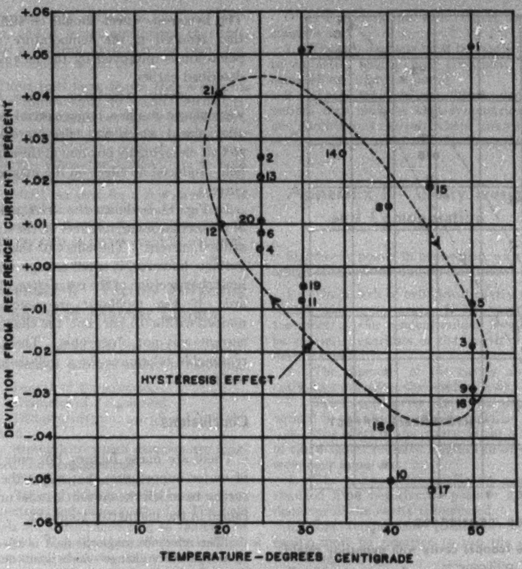


Fig. 8. Per-cent current deviation versus temperature (cold-rolled steel cavity)
Reference current = 1.00440 milliamperes

magnetic shield this difficulty was overcome.

A temperature-compensating nickel alloy was tried as a means for further reduction of the temperature coefficient. The permeability of this material decreases as the temperature increases. This may be used as a magnetic shunt to compensate a magnet in which the flux decreases with increase in temperature. A strip of the compensating alloy is shown attached to the magnet holders in Fig. 7. This strip is 0.02 inch by 1/8 inch by 5/8 inch, which yields a temperature coefficient close to zero. A strip 2 inches long causes the current to increase about 2 per cent for a 20-degree-centigrade temperature change. The compensating alloy supplies a convenient method of adjustment of temperature compensation.

If the mounting screws are tightened on the magnet ends, then, as the temperature changes, the pressure on the ends of the magnets will vary, causing a change in flux by magnetostrictive effect. To avoid this, the screws are backed by small nonmagnetic compresion springs.

Significant points in connection with cavity design and compensation are summarized in Appendix II.

In Fig. 8 is shown a plot of per-cent

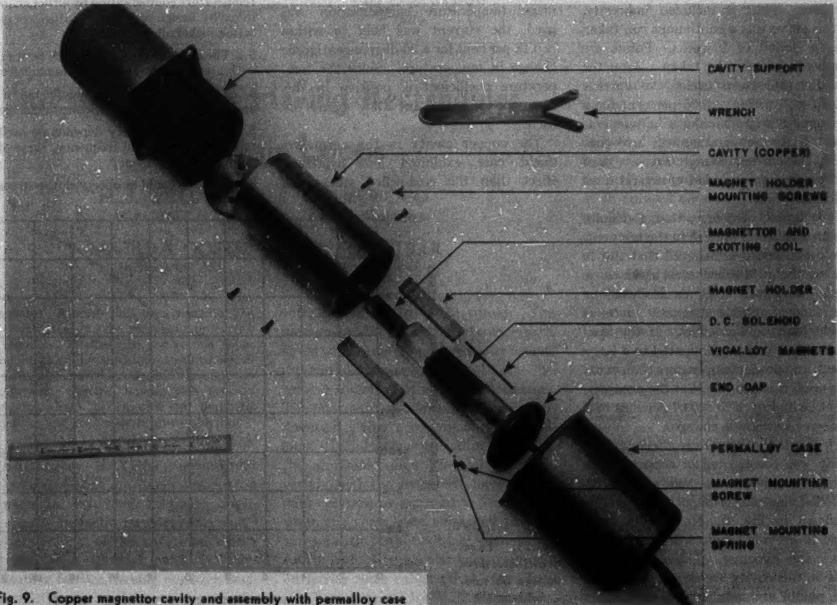

CAVITY SUPPORT
WRENCH
CAVITY (COPPER)
MAGNET HOLDER MOUNTING SCREWS
MAGNETTOR AND EXCITING COIL
MAGNET HOLDER
D. C. SOLENOID
VICALLOY MAGNETS
END CAP
PERMALLOY CASE
MAGNET MOUNTING SCREW
MAGNET MOUNTING SPRING

Fig. 9. Copper magnettor cavity and assembly with permalloy case

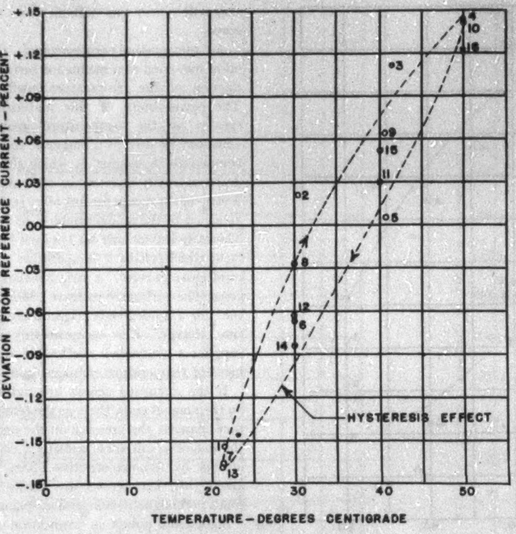


Fig. 10. Per-cent current deviation versus temperature (copper cavity with permalloy shield)
Reference current = 1.00380 milliamperes

change in current with respect to temperature. Time is included indirectly, as this curve was a continuous run taken over a period of 2 weeks. Points are numbered in the order taken. After the initial temperature cycle, the current stayed well within ±0.05 per cent for a 30-degree-centigrade change in temperature. If the temperature was held constant, the current change was less than 0.01 per cent for a period of several days at least.

It is noted, of course, that a definite hysteresis effect is present in the temperature cycle, and it is believed that this is due to the cold-rolled steel cavity. A change to silicon steel, Armco iron, or other more suitable magnetic material should improve this effect. If the hysteresis effect can be eliminated to a large extent, and proper temperature compensation used, it is possible that the accuracy might be improved to ±0.01 per cent to a 30-degree-centigrade change.

To check the hysteresis effect, a new cavity was constructed of copper and a permalloy shield case was used. The exploded view of this magnettor assembly is shown in Fig. 9. No attempt was made to provide temperature compensation in this structure. The results of the tests on this cavity are shown in Fig. 10. The points are numbered in the order taken. In spite of the fact that no auxiliary temperature compensation was used, the current was held to within ±0.18 per cent for a 30-degree-centigrade change in temperature. Since the temperature coefficient is positive in this case, the compensating alloy cannot be used.

The copper cavity in the permalloy shield case exhibited less hysteresis effect than the cold-rolled steel case. The hysteresis effect should be still further reduced if the temperature compensation is improved by the techniques described earlier.

The cavity structures during tests were placed inside a larger special permalloy shield which was effective enough so that changing its position in the earth's field produced no effect on the magnettor current.

In Fig. 11 is shown the drive voltage of the oscillator and its effect on the standardized current. The effect of the drive voltage, however, is small enough not to limit the precision of the magnettor standard. If the oscillator frequency remained within 0.1 per cent, the change in current was not observable. These are reasonable tolerances for a well-designed oscillator.

Conclusions

There are many factors, but only one of major importance, that affect the accuracy to which the current can be maintained in the magnettor standard.

1. The reference magnetic field is affected by temperature change which is, no doubt, the main source of error. All other effects are really secondary in nature.
2. The effect of temperature can be overcome to a satisfactory degree by proper temperature compensation by either changes in cavity design or by temperature-compensating nickel alloy.
3. The hysteresis effect needs to be studied with a view toward reducing its magnitude.
4. The magnettor core is unaffected in its operation by normal temperature changes.
5. While the accuracy depends on oscillator drive voltage and frequency, these offer no problem at present.
6. As far as could be observed, the strength

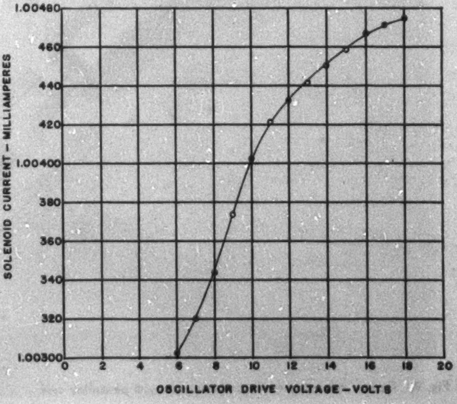


Fig. 11. Effect of oscillator drive voltage on solenoid current

of magnets with respect to time was very stable, and it is not anticipated that this will be a problem since they are used at a low flux density in the cavity.

7. These tests have shown that a current can be standardized to within ±0.05 per cent for a 30-degree-centigrade temperature change with the possibility of reducing it to ±0.01 per cent for the same change in temperature. When the temperature is held constant the present accuracy is better than ±0.01 per cent for a few days. It is conceivable that with temperature regulation and by careful design accuracies of ±0.01 per cent or better could be obtained.

Appendix I. Factors Affecting Choice of Magnet Length

The length of magnets for minimum quadrature voltage depends upon the following:

1. Length of d-c solenoid.
2. Diameter of d-c solenoid.
3. Distance magnets are placed from the solenoid.
4. Strength to which magnets are magnetized and demagnetized.
5. Composition of magnets.
6. Type of field in which magnets are magnetized and demagnetized. The pole kernels are shifted according to the type of field used. Considerable difference is noted, for example, when the magnets are demagnetized in a solenoid field or in the field between poles of an electromagnet.

Other considerations with respect to the magnets are:

1. The two magnets must be identical as to diameter, length, heat treatment, and composition (matched pairs).
2. When magnets are being demagnetized they must be kept symmetrical to one another and thereafter kept away from other magnetic materials and strong fields.

Appendix II. Cavity Design and Compensation

Significant points in connection with the design of the cavity and its treatment:

1. The length of cylindrical cavity is of considerable importance. If the cavity is too short, a poor residual nulling level will be obtained regardless of the length of the magnets used.
2. The diameter of the cavity apparently does not affect the residual level.
3. The steel cavity must be demagnetized in the absence of a d-c field.
4. The effect of the cavity on the length of the magnets required appears to be only secondary in nature.
5. The temperature coefficient can be changed from negative to positive as the diameter of the cavity is increased.
6. The temperature coefficient of the cavity must be negative to use the compensating alloy.
7. The amount of compensating alloy to use depends on placement of strip; strength to which a given set of magnets is magnetized; and heat treatment of magnets.

References

1. AIR-BORNE MAGNETOMETERS FOR SEARCH AND SURVEY, E. P. Felch, W. J. Means, T. Slonczewski, L. G. Parratt, L. H. Rumbaugh, A. J. Tickner. *AIEE Transactions*, vol. 66, 1947, pp. 641–51.

2. THE FUNDAMENTAL LIMITATIONS OF THE SECOND HARMONIC TYPE OF MAGNETIC MODULATION AS APPLIED TO THE AMPLIFICATION OF SMALL D.C. SIGNALS, F. C. Williams, S. W. Noble. *Proceedings*, Institution of Electrical Engineers, London, England, vol. 97, pt. II, Aug. 1950, p. 445.

3. A SATURATED CORE RECORDING MAGNETOMETER, D. C. Rose, J. N. Bloom. *Canadian Journal of Research*, Ottawa, Canada, vol. 28, Aug. 1950, pp. 153–63.

4. SOME DEVELOPMENTS IN ELECTRONIC MAGNETOMETERS, A. W. Brewer, J. Quires, H. McG. Ross. *Elliott Journal*, Elliott Brothers Limited, London, England, vol. I, no. 2, Sept. 1951.

5. MAGNETIC MODULATORS, E. P. Felch, V. E. Legg, F. G. Merrill. *Electronics*, New York, N. Y., vol. 25, no. 2, Feb. 1952, p. 113.

6. DETECTION SYSTEM, E. P. Felch, F. G. Merrill, T. Slonczewski. United States Patent No. *2,448,389*, Nov. 15, 1949.

7. DIRECT-CURRENT INDICATION, Cramer W. La Pierre. United States Patent No. *2,053,154*, Sept. 1, 1936.

8. STABILIZED PERMANENT MAGNETS, P. P. Cioffi. *AIEE Transactions*, vol. 67, pt. II, 1948, pp. 1540–43.

9. SPECIAL MAGNETIC AMPLIFIERS AND THEIR USE IN COMPUTING CIRCUITS, H. S. Sack, R. T. Beyer, G. H. Miller, J. W. Trischka. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 35, Nov. 1947, pp. 1375–82.

No Discussion

Stressed Ferrites Having Rectangular Hysteresis Loops

H. J. WILLIAMS
NONMEMBER AIEE

R. C. SHERWOOD
NONMEMBER AIEE

MATILDA GOERTZ
NONMEMBER AIEE

F. J. SCHNETTLER
NONMEMBER AIEE

Synopsis: A study has been made of the effect of stress on the magnetic properties of ferrites. Rectangular hysteresis loops were obtained by encasing toroidal specimens in plastics which shrink during polymerization. Ferrites having this type of hysteresis loop are useful in magnetic switching and magnetic memory devices.

FERRITES HAVING rectangular hysteresis loops are important for use in magnetic switching, magnetic amplifiers, and magnetic memory devices.[1-3] Hysteresis loops of this nature have been obtained with toroidal cores of polycrystalline ferrites, having negative magnetostriction, by applying pressure on the outer curved surface of the toroid. This pressure produces an internal compressive stress which acts circumferentially around the toroid causing this direction to become a direction of easy magnetization. The magnetic domains line up in this direction giving a rectangular hysteresis loop. If the magnetostriction is positive, the domains align themselves transversely and the hysteresis loop flattens out. Hysteresis loops have been traced while the cores were subjected to pressures ranging up to 1,200 pounds per square inch. Rectangular hysteresis loops on ferrite cores subjected to stress have also been obtained by W. N. Papian.[4]

A practicable method used for stressing the cores was to encase the wound cores in a plastic which shrinks when it polymerizes. Hysteresis loops were taken before and after encasing the cores in plastic and changes in the hysteresis loops of an encased core were observed as the temperature was varied from 0 to 66 degrees centigrade. The switching time of a core having a coercive force of 0.1 oersted was measured as a function of the applied field. Most of the ferrites have negative magnetostriction, so that this technique is applicable to a great variety

---

Paper 53-260, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 17, 1953; made available for printing April 20, 1953.

H. J. WILLIAMS, R. C. SHERWOOD, MATILDA GOERTZ, and F. J. SCHNETTLER are with the Bell Telephone Laboratories, Murray Hill, N. J.

The authors wish to thank S. O. Morgan, J. K. Galt, J. H. Rowen, C. L. Hogan, and C. D. Owens for helpful discussions, W. J Clarke for selecting a suitable plastic and casting the cores, J. G. Walker for handling the reproduction of the figures, and F. J. Dempsey for tracing loops on the Cioffi fluxmeter.

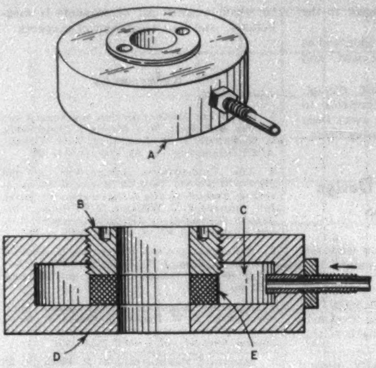


Fig. 1. Device for applying radial pressure to a toroidal specimen

- A. Assembled unit
- B. Plug
- C. Oil chamber
- D. Brass casing
- E. Specimen

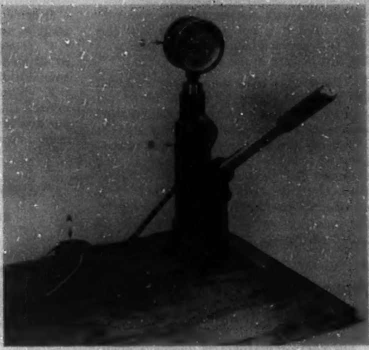


Fig. 2 (right). Complete apparatus for applying radial pressure

- A. Device shown in Fig. 1
- B. Hydraulic jack
- C. Oil pressure gauge

of compositions having a wide range in values of coercive force.

The stressing of ferrites may be found useful for other applications, since a compressive stress tends to establish an axis of easy magnetization parallel to the stress when the magnetostriction is negative, thus making a polycrystalline material assume the magnetic characteristics of a single crystal having one axis of easy magnetization. A tensile stress would have a similar effect for a material with positive magnetostriction. The effective anisotropy is a function of the stress and the magnetostriction.

Experimental Procedure

Toroidal specimens having an outside diameter of approximately 0.7 inch, an inside diameter of 0.4 inch, and an axial height of 0.15 inch were used. The upper and lower surfaces of the toroidal specimen were ground flat and parallel on a surface grinder, using a diamond wheel; then the specimen, Fig. 1, *E*, was placed inside the hollow toroidal brass casing *D*, and the threaded plug *B* screwed into place. Hydraulic pressure was applied to the outer curved surface of the toroid by forcing oil into the chamber *C*.

The complete assembly for applying hydraulic pressure is shown in Fig. 2. Pressure was applied by means of the hydraulic jack *B*, and the value of the pressure determined by a pressure gauge *C* mounted on the jack. Primary and secondary windings were wound through the toroidal assembly *A*, and hysteresis loops traced on a Cioffi recording fluxmeter as the pressure on the outer curved surface of the toroid was varied from 0 to 1,200 pounds per square inch.

The pressure *P* applied in this manner produces a compressive stress[5] in the core, which has a tangential component σ_θ and a radial component σ_r. σ_θ is a maximum at the inside diameter of the toroid, and has its minimum at the outside diameter; while σ_r is zero at the inside diameter, and has its maximum value at the outside diameter. However, σ_θ is always larger than σ_r, and it is the former component which determines the direction of easy magnetization. The average value of σ_θ is given by the equation

$$\sigma_\theta = bP/(b-a)$$

Table I. Chemical Composition and Change of Magnetic Constants with Pressure of Ni–Zn Ferrites

Specimen	Atomic Per Cent of Metal Content	Magnetic Constants	Radial Pressure, Pounds per Square Inch												
			0	100	200	300	400	500	600	700	800	900	1000	1100	1200
1758-5	66.39 Fe	B_m	1,700			1,900	1,900	1,950	1,950	1,955	1,975	1,985	1,915		2,000
	10.0 Ni	B_R	650			1,100	1,250	1,300	1,375	1,375	1,400	1,425	1,475		1,550
	23.5 Zn	B_R/B_m	0.38			0.58	0.66	0.66	0.70	0.75	0.71	0.72	0.74		0.77
		H_c	0.053			0.06	0.055	0.058	0.057	0.053	0.05	0.05	0.055		0.052
1785-11	66.67 Fe	B_m	1,725					1,825		1,850			1,875		
	10.0 Ni	B_R	475					1,500		1,555			1,570		
	23.33 Zn	B_R/B_m	0.27					0.82		0.84			0.84		
		H_c	0.021					0.025		0.025					
1579-36	67.00 Fe	B_m	1,530		1,650	1,650	1,650	1,650	1,730	1,725	1,775	1,775	1,775		1,775
	9.87 Ni	B_R	725		925	1,000	1,050	1,070	1,200	1,225	1,390	1,460	1,500		1,510
	23.04 Zn	B_R/B_m	0.470		0.56	0.60	0.64	0.65	0.68	0.70	0.78	0.82	0.84		0.85
		H_c	0.03		0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.032		0.04
1614-18	67.54 Fe	B_m	1,530		1,670	1,735	1,800	1,800	1,810	1,865	1,870	1,900	1,900		
	9.74 Ni	B_R	890		1,065	1,200	1,275	1,310	1,390	1,475	1,500	1,590	1,590		
	22.72 Zn	B_R/B_m	0.58		0.63	0.69	0.71	0.72	0.76	0.79	0.80	0.83	0.84		
		H_c	0.085		0.085	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08		
1796-5	68.60 Fe	B_m	2,100								2,230				
	9.42 Ni	B_R	1,670								1,370				
	21.98 Zn	B_R/B_m	0.79								0.61				
		H_c	0.12								0.19				

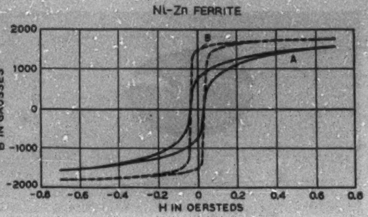


Fig. 3. Hysteresis loops for Ni-Zn ferrite specimen no. 1579-36 (67.09 per cent Fe)

A. Without pressure
B. With a pressure of 1,000 pounds per square inch

where b and a are the outside and inside diameters, respectively.

Magnetic Results With Pressure Device

Tests were made on a series of Ni—Zn ferrites. Specimen *1785-11* has the stoichiometric ratio, $Ni_{0.3}Zn_{0.7}Fe_2O_4$; the others have the same Ni to Zn ratio of 3 to 7, but with varying amounts of iron. The chemical compositions of the specimens in this series are given in Table I.

The effect on the hysteresis loop of applying a pressure of 1,000 pounds per square inch on specimen *1579-36*, in which 67.1 atomic per cent of the metal content is iron, is shown in Fig. 3. The remanence increased from 750 to 1,500 gausses, while the coercive force remained the same. Fig. 4 shows some of the magnetic properties of the above specimen:

1. $\mathbf{B}_m$, the value of the magnetization for a field **H** of 0.7 oersted;
2. $\mathbf{B}_R$, the remanence;
3. The ratio $\mathbf{B}_R/\mathbf{B}_m$;
4. The coercive force $\mathbf{H}_c$, plotted as a function of the pressure.

The ratio $\mathbf{B}_R/\mathbf{B}_m$ increases from a value of 0.46, for no pressure, to 0.85 for a pressure of 1,200 pounds per square inch. There is no appreciable change in the coercive force up to a pressure of 900 pounds per square inch, while from there on there is a slight upward trend. In order to determine the effect of higher stresses on the coercive force of a sample, the inside diameter of specimen *1614-17* was increased so that a given value of P would produce a value of σ_θ approximately three times as great as that obtained with the previously mentioned specimen. When subjected to a pressure of 1300 pounds per square inch, the coercive force of this specimen increased from 0.09 to 0.18 oersted.

Curves are shown in Fig. 5 for specimen *1785-11*, which has the stoichiometric ratio previously mentioned. Specimen *1796-5*, with 68.6 atomic per cent of its metal content iron, has the highest iron content of the series. This specimen showed a decrease in remanence and an increase in coercive force when pressure was applied, as seen in Table I. Harvey, Hegyi, and Leverenz[6] show that the magnetostriction changes from a negative to a positive value near this composition. The exact composition where this occurs, however, it not given. The compositions with higher iron content have positive and those of lower content have negative magnetostriction. It is quite likely that the composition of specimen *1796-5* lies on the positive side, which would explain why the remanence decreased with pressure. It should be mentioned here that, in general, the magnetic properties returned to their original values after the release of pressure.

Curves showing the permeability μ as a function of the flux density **B**, obtained with and without pressure, are given in Fig. 6 for specimen *1785-2*. The maximum permeability increased with pressure from 12,600 to 36,800. This is

Fig. 4. Effect of radial pressure on magnetic properties of Ni-Zn ferrite specimen no. 1579-36 (67.09 per cent Fe)

B_m = flux density for a field of 0.7 oersted
B_R = remanence
H_c = coercive force

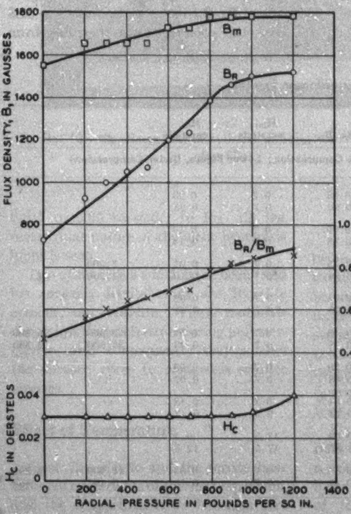
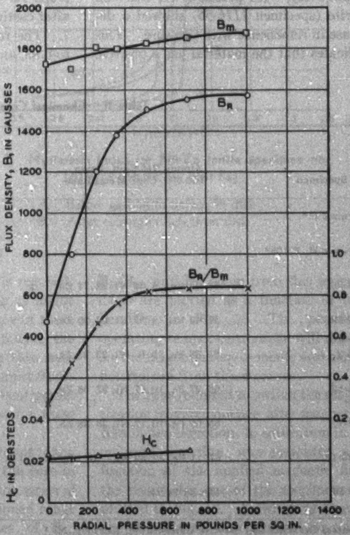


Fig. 5 (right). Effect of radial pressure on magnetic properties of Ni-Zn ferrite specimen no. 1785-11 (66.67 per cent Fe)

B_m = flux density for a field of 0.7 oersted
B_R = remanence
H_c = coercive force



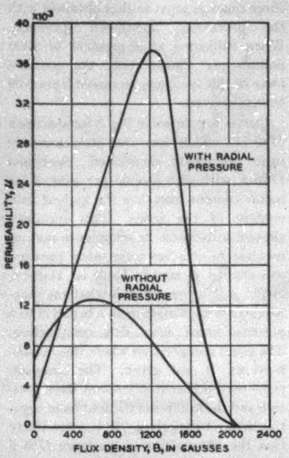
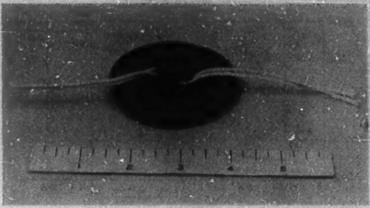


Fig. 6. Permeability versus flux density curves obtained with and without radial pressure on Ni–Zn ferrite specimen no. 1785-2 (66.67 per cent Fe)

the highest maximum permeability thus far reported for a polycrystalline ferrite.

Table II gives data showing the effect of pressure on a number of different ferrites. Specimens of Ferramic *G* and *H* with $\mathbf{B}_R/\mathbf{B}_m$ ratios 0.63 and 0.69 without pressure show increases to 0.90 and 0.88, respectively, with pressure. A Mn—Zn ferrite (specimen *1116-56*) showed a decrease in remanence with pressure. This indicates that the material has a positive magnetostriction. However, for a material with positive magnetostriction, tension should cause the remanence to increase. Further tests are being made along this line.

Fig. 7 (right). Wound toroidal specimen cast in Selectron disk



Cores Cast in Plastic

Another method used to stress a specimen in order to obtain a rectangular hysteresis loop was to encase a core, with its primary and secondary windings, in a rigid grade of casting resin. For this purpose, a hard type of styrene polyester resin was employed, namely Selectron *5003*. During polymerization, this resin shrinks about eight per cent by volume, thus exerting a considerable stress on the specimen. Fig. 7 shows a wound core cast in a Selectron disk having a thickness only slightly greater than the thickness of the core. With this method a whole array of cores could be cast in resin at one time. Fig. 8 shows hysteresis loops for specimen *1785-10*, before and after casting in a Selectron disk; see Fig. 7. The ratio $\mathbf{B}_R\mathbf{B}_m$ increased from 0.34 to 0.83, and the coercive force from 0.017 to 0.05 oersted. As the maximum field strength was gradually decreased, loops were traced on the above mentioned specimen; see Fig. 9.

To check the effect of thickness of the plastic encasement on the magnetic properties a core of similar composition, specimen *1785-19* was cast in the center of a cylinder of plastic as shown in Fig. 10. None of the cast disks of Selectron showed any signs of cracking, but a large cylinder of Selectron developed cracks. To avoid this condition a silica and glass fiber-filled Selectron was used in the casting of this cylinder. The filled compound shrinks only about 6 per cent by volume, but this is apparently enough to impart an appreciable stress on the ferrite core.

Fig. 11 shows the hysteresis loops obtained on the specimen shown in Fig. 10, before and after casting. In this case, the coercive force increased from 0.02 to 0.13 oersted, which is a much greater increase than is observed when a core is cast in a disk. This can be seen by comparing these values with those in Fig. 8, where $\mathbf{B}_R/\mathbf{B}_m$ increases from 0.43 to 0.88 and the coercive force increases from

Table II. Chemical Composition and Effect of Compression on Ferrites

Specimen	Atomic Per Cent, Using 100-Per-Cent Metal Base	B_R, gausses	B_m, gausses	B_R/B_m	H_m, oersteds	H_c, oersteds	μ_m	μ_0
		(Upper Figure, No Compression; Lower Figure, Under Compression)						
Ferramic G*		1,040 2,350	1,640 2,610	0.63 0.90	0.7 0.7	0.23 0.23		
Ferramic H, F 108*		1,450 2,320	2,100 2,630	0.69 0.88	0.7 0.7	0.15 0.14		
1614-20	67.54 Fe, 9.69 Ni, 22.77 Zn	900 1,630	1,640 1,930	0.55 0.84	0.7 0.7	0.07 0.10	7,900 12,300	
1579-19	67.09 Fe, 9.82 Ni, 23.09 Zn	730 1,250	1,700 1,780	0.43 0.70	0.7 0.7	0.02 0.04		
1785-2	66.67 Fe, 10.0 Ni, 23.33 Zn	320 1,460	1,720 1,870	0.18 0.78	0.7 0.7	0.02 0.03	12,600 36,800	6,500 2,500
1615-12	66.27 Fe, 10.07 Ni, 23.66 Zn	470 1,260	1,700 1,850	0.28 0.68	0.7 0.7	0.03 0.07		
1759-1	65.95 Fe, 10.17 Ni, 23.88 Zn	460 1,370	1,350 1,880	0.34 0.73	0.7 0.7	0.13 0.11		
S1525-12	67.11 Fe, 32.89 Ni, 0 Zn	820 1,730	1,400 1,980	0.59 0.87	17.5 17.5	9.5 12.5		
1116-56	Mn-Zn Ferrite	750 550	2,190 1,330	0.34 0.41	0.7 0.7	0.03 0.10	12,400 6,100	2,800 800

* General Ceramics and Steatite Corporation, Keasbey, N. J.

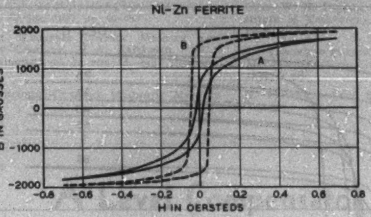


Fig. 8. Hysteresis loops for Ni-Zn ferrite specimen no. 1785-10 (66.67 per cent Fe)

A. Before casting in Selectron disk of Fig. 7
B. After casting in Selectron disk of Fig. 7

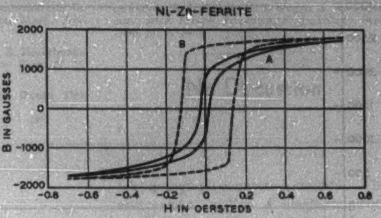


Fig. 11. Hysteresis loops for Ni-Zn ferrite specimen no. 1785-19

A. Before casting in Selectron cylinder of Fig. 10
B. After casting in Selectron cylinder of Fig. 10

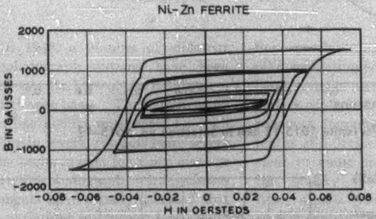


Fig. 9. Hysteresis loops for Ni-Zn ferrite specimen no. 1785-10 (66.67 per cent Fe) cast in Selectron

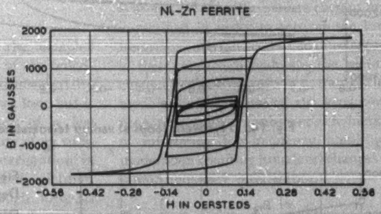


Fig. 12. Family of hysteresis loops for Ni-Zn ferrite specimen no. 1785-19 (66.67 per cent Fe) cast in Selectron cylinder of Fig. 10

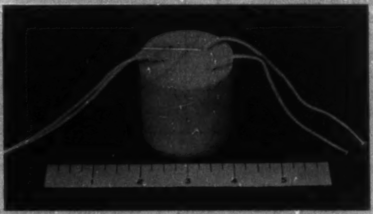


Fig. 10. Specimen cast in cylinder of Selectron with Fiberglass filler

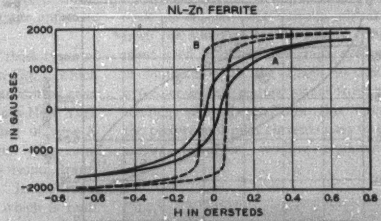


Fig. 13. Hysteresis loops for Ni-Zn ferrite specimen no. 1615-12 (66.27 per cent Fe)

A. Before casting in Selectron disk
B. After casting in Selectron disk

0.017 to 0.05 oersted. In Fig. 12, the rectangular nature of the inner hysteresis loops is seen.

Two other resins which have been tried for encasing ferrite cores are Bakelite epoxide and Araldite. These materials shrink less than Selectron during polymerization, and consequently they subject the encased cores to somewhat smaller stresses.

Effect of Temperature

Since the plastic and the ferrite have different coefficients of thermal expansion, and the magnetostriction becomes zero before the Curie temperature is reached, variations in temperature such as might be encountered in the use of one of these units might be sufficient to relieve the stress on the core. To check this, hysteresis loops were traced at temperatures ranging from 0 to 66 degrees centigrade. Specimen *1615-12* was selected for these tests.

Fig. 13 shows the hysteresis loops taken both before and after casting the core in a Selectron disk and Fig. 14, the series of loops for the previously mentioned range of temperatures. The values of $\mathbf{B}_m$, $\mathbf{B}_R/\mathbf{B}_m$, and $\mathbf{H}_c$ taken from this series of loops are plotted as a function of temperature in Fig. 15. The permitted operating temperature range will depend on the magnetic requirements and can be determined from these curves.

The time required to switch the magnetization from remanence, after magnetization in one direction, to saturation in the opposite direction, was measured as a function of the applied magnetic field; the results for one of the specimens are plotted in Fig. 16. A switching time of one microsecond was obtained for a field

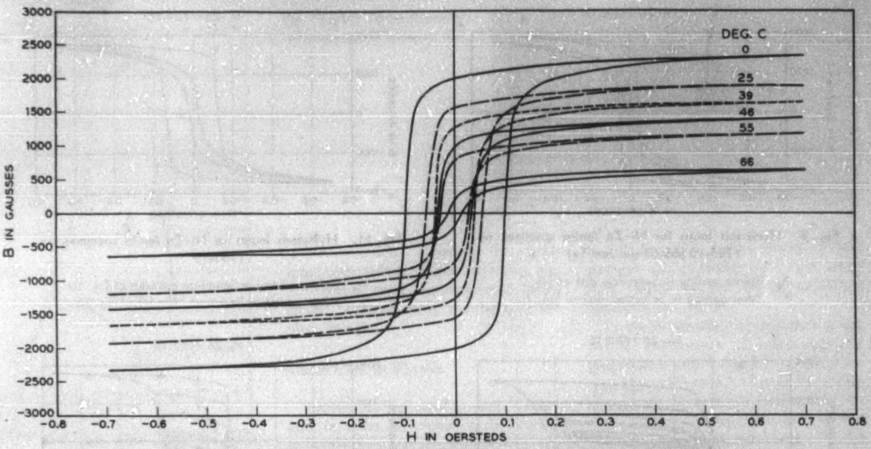


Fig. 14. Hysteresis loops at various temperatures for Ni–Zn ferrite 1615-12 cast in Selectron no. 1615–12

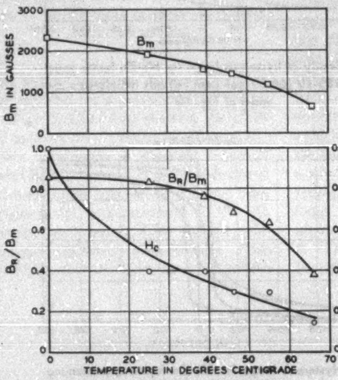


Fig. 15 (left). Dependence of magnetic properties on temperature for Ni–Zn ferrite specimen no. 1615-12 (66.27 per cent Fe) cast in Selectron disk

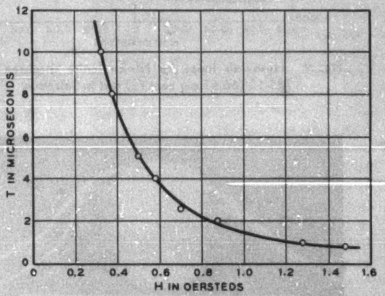


Fig. 16. Switching time as a function of the applied field for a specimen having a composition approximately $Ni_{0.3}Zn_{0.7}Fe_2O_4$

of approximately 1.3 oersteds. The core had a coercive force of 0.1 oersted after being encased in a plastic disk. The outside diameter of the core was 0.175 inch, the inside diameter 0.092 inch, and the thickness 0.040 inch.

Summary

Pressure applied on the external curved surface of a toroidal core of ferrite having a negative magnetostriction tends to establish a direction of easy magnetization which extends circumferentially around the core. The pressure increases the maximum permeability and the remanence, thus causing the hysteresis loop to become more nearly rectangular. Quantitative measurements of this effect were made on a series of specially prepared ferrites having various compositions, one of which had the stoichiometric ratio $Ni_{0.3}Zn_{0.7}Fe_2O_4$ and the others, the same Ni-to-Zn ratio, but with either slightly less or slightly more iron. Rectangular hysteresis loops with very low values of coercive force were obtained. This combination of magnetic properties should make the material very desirable for use in magnetic memory devices.

It has been shown that a simple and practicable means of stressing the cores to obtain rectangular hysteresis loops is to cast the wound cores in a resin which shrinks during polymerization. Although with this technique it is not yet certain that the stress will remain constant with time, one core cast in a Selectron disk has shown no change in magnetic properties after a period of 6 months, and others have shown no change for the shorter periods of time for which they have been observed. Further work on this aspect of the problem has been planned, which will involve testing at periodic intervals, cores cast in different plastics and then held at various temperatures.

References

1. Magnetic Storage and Delay Line, A. Wang. *Journal of Applied Physics*, New York, N. Y., vol. 21, Jan. 1950, pp. 49–54.

2. Digital Information Storage in Three Dimensions Using Magnetic Cores, J. W.

Forrester. *Journal of Applied Physics*, New York, N. Y., vol. 22, Jan. 1951, pp. 44–48.

3. STATIC MAGNETIC MATRIX MEMORY AND SWITCHING CIRCUITS, J. A. Rajchman. *RCA Review*, Princeton, N. J., vol. 13, June 1952, pp. 183–201.

4. A COINCIDENT-CURRENT MAGNETIC MEMORY UNIT, W. N. Papian. *Thesis*, Massachusetts Institute of Technology, Cambridge, Mass., Aug. 1950.

5. THEORY OF ELASTICITY (book), S. Timoshenko. McGraw-Hill Book Company, New York, N. Y., 1934, p. 56.

6. FERROMAGNETIC SPINELS FOR RADIO FREQUENCIES, R. L. Harvey, I. J. Hegyi, H. W. Leverenz. *RCA Review*, Princeton, N. J., vol. 11, Sept. 1950, pp. 321–63.

◆

No Discussion

A Level Compensator for Telephotograph Systems

T. A. JONES
NONMEMBER AIEE

W. A. PHELPS
NONMEMBER AIEE

SINCE 1936, a number of different telephotograph systems[1] have been operated over the lines of the Bell System on a leased-wire basis. The requirements which a line facility must meet for such service are in several respects more severe than the requirements for voice-message service. Telephotograph - transmission requirements have usually specified that abrupt variations in line net loss should be limited to ±0.2 decibel, single frequency noise power should be at least 50 decibels below maximum signal power, and throughout the useful frequency band (1,200 to 2,600 cycles per second), the envelope delay should be equalized within ±300 microseconds. Such requirements were originally satisfied by using 4-wire *H-44-25* side circuits in cable, where available, and elsewhere by using 2-wire open-wire side circuits. These facilities were delay-equalized over the necessary frequency range, and precautions were taken to minimize various transmission disturbances.

However, the long voice-frequency loaded cable circuits, such as those which formerly connected New York and Chicago, have been largely supplanted by multichannel (so-called broad-band) type-*K* carrier telephone systems[2,3] operating on nonloaded cable pairs, or by type-*L* carrier telephone systems[4] working over coaxial cables. Special services such as carrier telegraph, program, and telephotograph transmission must now use channels of these broad-band carrier systems for long-haul applications, and this requirement has brought new problems. For speech transmission, the complex pilot regulating equipment of the broad-band systems accomplishes substantially perfect control of the line net loss, but it produces frequent variations of a few tenths of a decibel, which are often sufficient to cause marked deterioration in the quality of transmitted pictures. Also, signals passing through type-*K2* carrier systems are slightly affected by 60- and 120-cycle-per-second modulation caused by the a-c filament supply used for some of the amplifier tubes. This effect also is negligible for speech transmission, but is sometimes quite apparent in transmitted pictures.

The possible severity of these types of interference is illustrated in Fig. 1, which is a section of a picture transmitted over a New York-Chicago-Atlanta-New York loop composed of a number of type-*K* carrier channels in tandem. For comparison, the same picture before transmission is shown in Fig. 2. In the transmitted picture of Fig. 1, which is built up of vertical lines, it may be noted that in some areas which should have a constant shade almost every line has a different density, and on each side of the picture there is a group of very dark lines caused by larger level changes. The whole picture is covered by slanting striations due to power frequency interference. It should be realized, however, that these defects are normally less pronounced than in Fig. 1, which was chosen to illustrate the problem clearly.

Instead of modifying the carrier equipment to eliminate the interference at its several sources, it was deemed more practical to cancel the interference from the signal delivered by the carrier facility. This function is performed by the recently developed telephotograph level compensator, which consists of a pilot channel arrangement designed for insertion in the telephotograph connecting circuits. This arrangement utilizes a single-frequency pilot current, which is transmitted over the line facility along with the picture current, and is modulated during transmission by the same interference which modulates the picture signal. After transmission the modulated pilot current is separated from the picture current and is rectified in order to recover the low-frequency interference. The interference is then used to remodulate the picture current in phase opposition to the existing modulation. The use of the level compensator results in a substantial reduction of interfering low-frequency amplitude modulation resulting from level changes or power-frequency interference.

Description of the Level-Compensator Terminal Circuits

The level-compensator sending circuit is shown in Fig. 3. In this circuit a 3,145-cycle-per-second pilot current, obtained from the standard carrier supply equipment used in carrier telegraph systems, is combined with the 1,200- to 2,600-cycle-per-second picture signal from the sending subscriber. From this circuit the picture and pilot currents pass to the carrier telephone sending equipment.

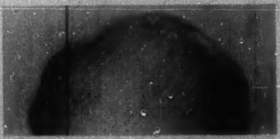


Fig. 1. Variations in shade produced by pilot regulator action and power-frequency interference on a multisection type-K carrier circuit

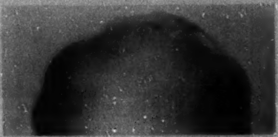
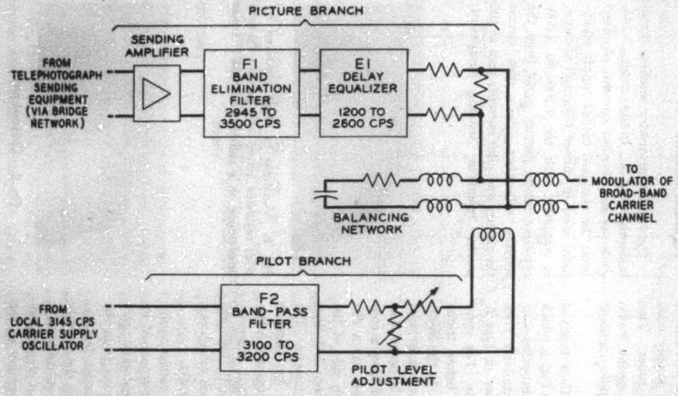
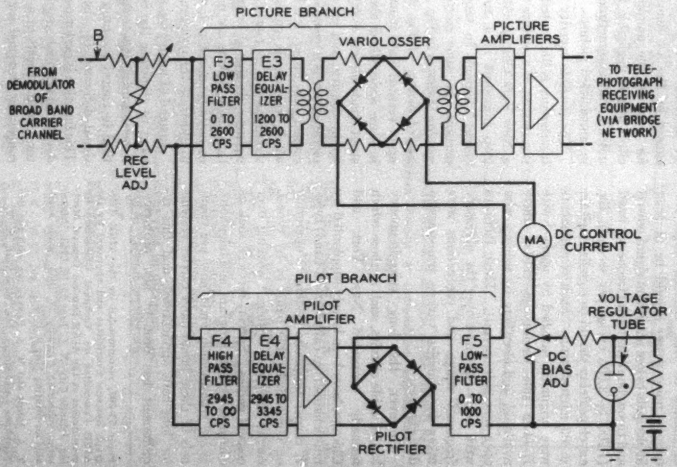


Fig. 2. Original print

Paper 53-297, recommended by the AIEE Telegraph Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted February 17, 1953; made available for printing May 5, 1953.

T. A. JONES and W. A. PHELPS are with the Bell Telephone Laboratories, Inc., New York, N. Y.


PICTURE BRANCH
SENDING AMPLIFIER
FROM TELEPHOTOGRAPH SENDING EQUIPMENT (VIA BRIDGE NETWORK)
F1 BAND ELIMINATION FILTER 2945 TO 3500 CPS
E1 DELAY EQUALIZER 1200 TO 2600 CPS
BALANCING NETWORK
TO MODULATOR OF BROAD-BAND CARRIER CHANNEL
PILOT BRANCH
FROM LOCAL 3145 CPS CARRIER SUPPLY OSCILLATOR
F2 BAND-PASS FILTER 3100 TO 3200 CPS
PILOT LEVEL ADJUSTMENT


PICTURE BRANCH
B
FROM DEMODULATOR OF BROAD BAND CARRIER CHANNEL
REC LEVEL ADJ
F3 LOW PASS FILTER 0 TO 2600 CPS
E3 DELAY EQUALIZER 1200 TO 2600 CPS
VARIOLOSSER
PICTURE AMPLIFIERS
TO TELEPHOTOGRAPH RECEIVING EQUIPMENT (VIA BRIDGE NETWORK)
MA
DC CONTROL CURRENT
PILOT BRANCH
F4 HIGH PASS FILTER 2945 TO ∞ CPS
E4 DELAY EQUALIZER 2945 TO 3345 CPS
PILOT AMPLIFIER
PILOT RECTIFIER
F5 LOW-PASS FILTER 0 TO 1000 CPS
DC BIAS ADJ
VOLTAGE REGULATOR TUBE

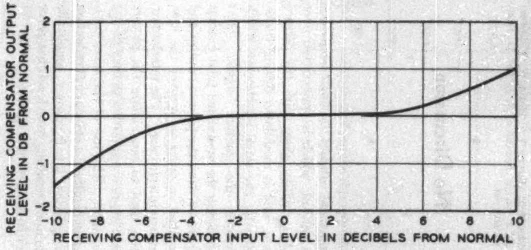
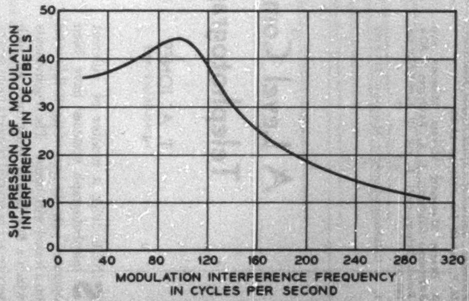


Fig. 3 (left). Level-compensator sending circuit

Fig. 4 (lower left). Level-compensator receiving circuit

Fig. 5 (above). Variation of output level with input level of level-compensator receiving circuit

Fig. 6 (below). Suppression of line-modulation interference by level compensator as a function of modulation frequency



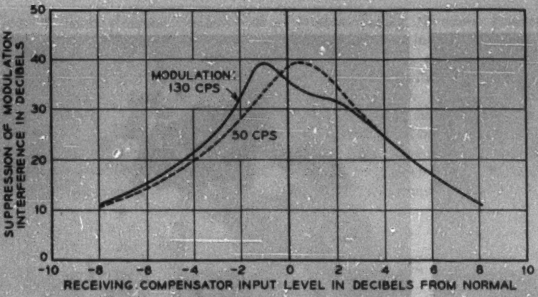
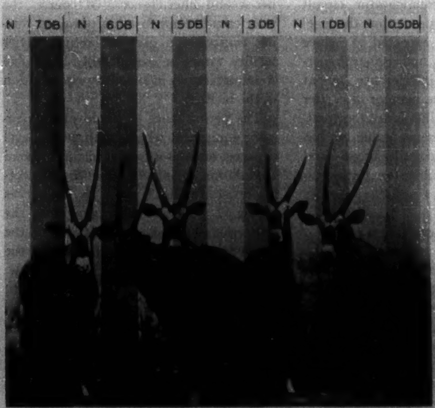


Fig. 7. Suppression of line-modulation interference by level compensator as a function of level at input of receiving compensator circuit

The pilot frequency has been chosen above rather than below the picture-signal band, so that the full frequency range of the carrier telephone channel below 2,700 cycles per second is available to the subscribers for alternate speech transmission, when desired. Filter *F1* provides substantial attenuation to picture or speech-frequency components lying in the pilot-channel band between 2,945 cycles per second and the 3,500-cycle-per-second upper limit of the carrier telephone channel. Such components are thereby prevented from actuating the receiving compensator equipment. Equalizer *E1* equalizes the delay of this filter over the picture signal frequency band. Filter *F2* attenuates certain extraneous frequency components which are present in the 3,145-cycle-per-second carrier supply. The level of the pilot current delivered by the sending circuit is adjusted 13 decibels below that of the maximum picture current. This adjustment permits the pilot current to be transmitted well above line noise, without causing appreciable intermodulation with the picture current.

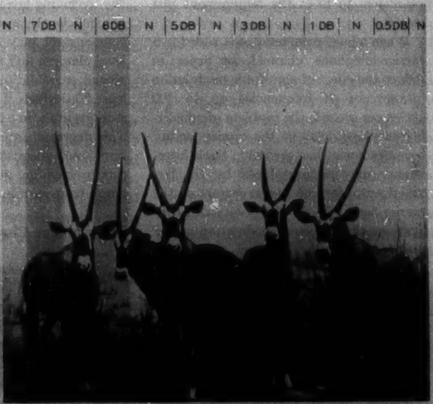
The level compensator receiving circuit, shown in Fig. 4, is connected between the output terminals of the broad-band carrier telephone channel and the receiving subscriber. In this circuit, the picture and pilot currents, together with any modulated interference picked up during transmission, are separated by low-pass and high-pass filters. The picture current from low-pass filter *F3* flows through delay-equalizer *E3*, which corrects the delay distortion of the filter, into a variolosser network consisting of a special type of balanced modulator. This network contains fixed resistors as series elements and a varistor bridge as a shunt element. The picture current flows through the variolosser at levels low enough to result in negligible signal distortion. The pilot current, after passing through high-pass filter *F4* and its delay equalizer *E4*, is amplified, rectified, and passed through low-pass filter *F5* in order to recover any low-frequency modulation to which the pilot current may have been subjected during transmission. This low-frequency current, in series with an adjustable d-c bias current, is used to control the resistance of the varistor bridge network in the variolosser, and thus to control the transmission loss introduced by the variolosser in the picture branch.

For effective operation of the level compensator the transmission loss of the variolosser, as controlled by the modulation envelope superimposed on the pilot current, must vary in magnitude and phase so as to cancel the interference envelope superimposed on the picture current. The magnitude requirement is met by proper adjustment of the pilot amplifier gain and d-c bias current. The phase requirement is met by designing the picture and pilot branches so that modulation frequencies up to about 200 cycles per second suffer substantially the same delay in the two branches.



Courtesy of the American Museum of Natural History

Fig. 8. Effect of level changes made in local test without level compensator



Courtesy of the American Museum of Natural History

Fig. 9. Effect of level changes made in local test with level compensator in circuit

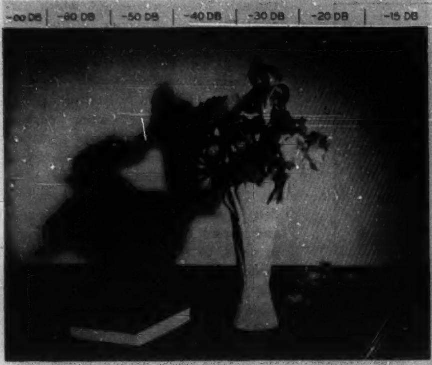


Fig. 10. Effect of 120-cycle-per-second modulation interference in local test without level compensator

-∞ DB
-60 DB
-50 DB
-40 DB
-30 DB
-20 DB
-15 DB

Fig. 11. Effect of 120-cycle-per-second modulation interference in local test with level compensator

Delay Equalization of the Carrier-Telephone Channel

As previously mentioned, for satisfactory operation of the types of telephotograph system under consideration, the maximum deviation of the envelope delay of the transmission circuit from end to end should not much exceed ±300 microseconds over the picture-signal frequency band from 1,200 to 2,600 cycles per second. When the picture signals are transmitted over a broad band carrier facility, the principal sources of delay distortion are the channel band-pass filters of the carrier-telephone terminals. This distortion must be corrected by delay equalizers to meet the requirement just stated.

When a level compensator is added to a carrier-telephone channel, in order to reduce the effect of amplitude modulation interference at frequencies up to 200 cycles per second, the problem of equalizing envelope delay of the carrier facility becomes somewhat greater. Such interference may be introduced by the terminal amplifiers or line repeaters of the carrier facility. At the points of introduction, simultaneous modulation of the picture and pilot carriers occurs. In order that the pilot-carrier modulation may be used effectively to cancel the picture-carrier modulation in the level-compensator receiving circuit, both must reach the variolosser in this circuit simultaneously. To accomplish this, the envelope delay effective over a frequency band, extending 200 cycles per second above and below the pilot frequency, must be adjusted to the same value as that effective over a band 200 cycles per second above and below the picture-carrier frequency. For this reason, the delay equalization of each carrier-channel filter carrying the picture and pilot currents must be effective over the pilot band as well as the picture band; that is, from 1,200 to 3,345 cycles per second. In a multisection carrier facility, consisting of channels of two or more carrier-telephone systems connected in tandem, the channel filters at each connecting point must be delay equalized at that point, so that interference introduced in any or all sections can be compensated at the receiving terminal.

The delay equalizers provided for the carrier-telephone channel filters and for the receiving-level compensator circuit are sufficiently accurate so that the level compensator may be expected to reduce level changes and 60- and 120-cycle-per-second modulation interference by at least 20 decibels if the number of carrier sections in the line facility is not too great. This degree of interference reduction has been found adequate to render broad-band carrier facilities suitable for telephotograph transmission.

Performance Characteristics

A typical characteristic showing the variation of the output level with the input level of a properly adjusted level compensator receiving circuit is shown in Fig. 5. This curve indicates that, in the absence of power frequency modulation, level changes in the line facility as large as ±5 decibels can be tolerated without exceeding the usual limit of ±0.2 decibel in the receiving subscriber's loop. The adjustments necessary to obtain this performance are easily made with the aid of simple testing equipment, and are stable over a period of several weeks.

After the circuit has been adjusted for level compensation, its typical performance with respect to modulation interference over the frequency range from zero to 300 cycles per second is as shown in Fig. 6. This curve applies to the level compensator sending and receiving circuits connected together by an artificial resistance line containing a source of modulation interference. The reduction of such interference due to the use of the compensator is plotted as a function of the modulation frequency, and amounts to nearly 40 decibels for 60 and 120-cycle-per-second interference. Of course, the interference compensation obtained during transmission over an actual carrier facility would be considerably lower due to imperfections in the delay equalization of the facility.

The performance of the level compensator equipment when simultaneous level changes and 50- or 130-cycle-per-second modulation are introduced in the connecting resistance line is given by the curves of Fig. 7. These curves show that the protection against modulation in the power-frequency range falls off as the line loss departs from its normal value. However, it is expected that a substantial compensation for such interference will be obtained in actual practice since the variations in line net loss normally encountered in broad-band carrier channels are small.

A sample picture, locally transmitted, containing line-level changes from 0.5 decibel to 7 decibels, with no level compensator in circuit, is shown in Fig. 8.

The effect of a level change of 0.5 decibel may be clearly seen in the picture. The same test picture transmitted with the level compensator in circuit is shown in Fig. 9. In this picture no change smaller than 6 decibels produces a noticeable effect.

Sample pictures were also locally transmitted to illustrate the performance of the level compensator in reducing the effect of 120-cycle-per-second modulation interference on the line. The picture of Fig. 10 was transmitted without the level compensator, and that of Fig. 11 was transmitted after the compensator had been added. No interference pattern is visible in Fig. 11, even for modulation levels only 15 decibels below the maximum picture level.

The tests which have so far been made over actual broad-band carrier facilities indicate that the level compensator should greatly extend their usefulness for telephotograph transmission.

References

1. A New Telephotograph System, **F. W. Reynolds.** *Bell System Technical Journal*, New York, N. Y., vol. 15, Oct. 1936, pp. 549–74.

2. A Carrier Telephone System for Toll Cables, **C. W. Green, E. I. Green.** *Bell System Technical Journal*, New York, N. Y., vol. 17, Jan. 1938, pp. 80–105.

3. Cable Carrier Telephone Terminals, **R. W. Chesnut, L. M. Ilgenfritz, A. Kenner.** *Bell System Technical Journal*, New York, N. Y., vol. 17, Jan. 1938, pp. 106–24.

4. Frequency Division Techniques for a Coaxial Cable Network, **R. E. Crane, J. T. Dixon, G. H. Huber.** *AIEE Transactions*, vol. 66, 1947, pp. 1451–59.

◆

No Discussion

L3 Coaxial System—Television Terminals

JOHN W. RIEKE
ASSOCIATE MEMBER AIEE

R. S. GRAHAM
ASSOCIATE MEMBER AIEE

Synopsis: Television terminals are required at circuit ends of the *L3* coaxial system, at the transmitting end to condition video signals for carrier transmission, and at the receiving end to detect the transmitted signals. Special signal characteristics, that is, a degree of modulation which exceeds the value commonly referred to as 100 per cent modulation, require departures from standard modulating and detecting processes. The high degree of modulation requires both careful control of transmitted wave form and, at the receiver product, demodulation with phase-synchronous carrier (homodyne detection).

Carrier regeneration requirements result in the choice of 1-step frequency translation from the video frequency spectrum to the allocated vestigial side-band carrier spectrum. The 1-step process using a single modulator results in unusual balance requirements for the modulator itself and an unusual circuit configuration.

Transmission quality objectives for the terminals are such as to permit six pairs of television (TV) terminals to operate in tandem in a transcontinental circuit. This permits a degree of interconnecting flexibility in operation with other systems, that is, *L1* coaxial or microwave systems. These objectives place severe requirements upon the transmission stability of various filters and other circuits within the terminals. New network techniques both in design and fabrication are brought to bear in the effort to achieve required performance. The transmitting and receiving terminals are described, illustrating the functional operation and mechanical and electrical arrangements of the equipment.

THE MAIN features of the *L3* coaxial cable transmission system, including its TV transmitting and receiving terminal equipment have been described.[1] The transmitting terminal conditions a TV signal for carrier transmission over the system simultaneously with a group of 600 telephone messages. The receiving terminal reconverts the carrier signal at each receiving point along the cable route. Primarily, the transmitting terminal is a modulator which translates the composite video picture spectrum of frequencies up to the carrier band of frequencies, and the receiving terminal is a detector which retranslates the carrier spectrum back to its original band of frequencies.

Particular characteristics of the transmitted TV signal, which are intended to aid in achieving optimum transmission quality, have necessitated the departures from past techniques in modulation and demodulation processes that are described in the following. Described also are the methods employed to achieve transmitted picture quality adequate for tandem operation of as many as six pairs of transmitting and receiving terminal equipments in a 4,000-mile TV transmission circuit. Operation with several pairs of terminals in tandem occurs when *L3* coaxial systems are interconnected with *L1* coaxial systems or microwave radio systems.

L3 TV terminal development has been in progress since early in 1948. Two transmitting and two receiving terminals have been built on a preproduction basis, and are currently being tested under field conditions as part of the *L3* system field trial. Development effort is continuing on the terminals with emphasis on equipment reliability, including means for maintaining and improving transmission quality.

Frequency Allocations

The *L3* coaxial system was designed to have as broad a transmission band as the economics of repeater spacing, together with presently realizable feedback amplifier performance, permit.[2] The band extends from 300 kc to 8.5 megacycles. In comparison with this band, a broadcast TV signal occupies the frequency spectrum from zero frequency up to 4.5 megacycles.

From the foregoing it is evident that the TV spectrum will not occupy fully the available system transmission band. It is feasible and attractive to allocate part of the transmission band for TV transmission, and the remainder for transmission of message channels. Detailed allocations then result from a compromise among transmission performance, cost, and the number of message channels made available.

From these considerations vestigial side-band transmission of the TV signal, rather than double side-band transmission is called for. The smaller the vestigial band of transmitted frequencies, the smaller will be the total TV band required. However, both the cost of band-shaping filters and the difficulty of maintaining satisfactorily low values of vestigial side-band quadrature distortion in-

---

Paper 53-224, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 18, 1953; made available for printing April 24, 1953.

John W. Rieke and R. S. Graham are with the Bell Telephone Laboratories, Inc., Murray Hill, N. J.

Acknowledgment is made to our associates for their many contributions including ideas relating to transmission processes and in the development of new circuit components which were necessary to the achievement of the quality and reliability objectives.

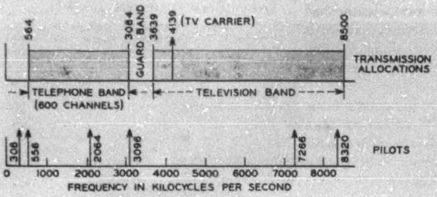


Fig. 1. Frequency allocations for L3 combined television-telephone transmission

crease as the vestigial bandwidth is reduced. The compromise of these factors resulted in the choice of a 500-kc vestigial side band.

Another choice made was to transmit the TV signal in the upper part of the *L3* band, and the message channels in the lower part. This allocation was determined by considering the noise distribution in the transmission band, together with the modulation distortion, (harmonic distortion), produced by the repeaters. By transmitting TV in the upper part of the band, a minimum modulation distortion is achieved since the harmonics of the TV signal largely fall outside the transmitted band, or at high frequencies where their effects in the picture are relatively less visible than low-frequency distortion. This factor outweighs the higher noise level in the upper part of the band.

As to the TV carrier location, it is placed at the bottom of the TV band at 4.139 megacycles, with the vestigial side band extending down to 3.64 megacycles and the main side band extending upward to 8.50 megacycles. Alternatively, the carrier could have been located at the top of the *L3* band with a main lower side band and vestigial upper side band, but this choice would be disadvantageous because of higher noise levels and poorer repeater gain stability near the top edge of the transmitted band.

The final allocation of frequencies is shown in Fig. 1. Just below the TV band from about 3 to 3.5 megacycles is a dead space. This frequency space is needed for the filters which are employed to separate the telephone signals from the TV signals at TV and telephone dropping points. A very high value of loss is required of these filters in their attenuating bands, and if this is built up over too small a frequency band the resulting delay distortion introduced into the TV band becomes very difficult and expensive to equalize. Below this guard band is the master group of 600 telephone channels, which is transmitted simultaneously with the TV signal.

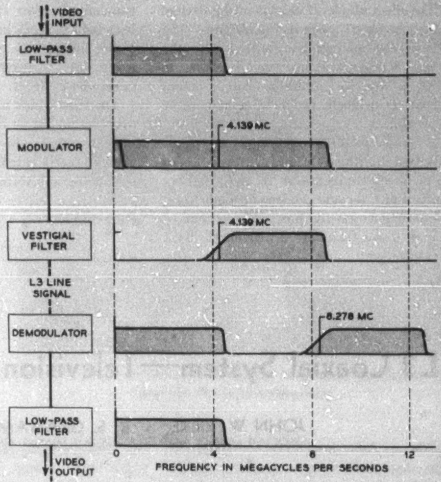


Fig. 2 (right). Television terminal modulation processes

The detailed allocation of specific frequencies, that is, TV carrier and the pilot frequencies, results from consideration of effects produced by these frequencies in telephone channels as a consequence of modulation distortion in repeaters. These considerations have been described in detail.[1]

Modulation Process

In the *L1* coaxial cable system[3] frequency translation by the TV terminals is accomplished in two stages. The 2-step process employs a first modulator supplied with a very-high-frequency carrier to translate all video frequencies to a band far outside the final transmitted band of frequencies where the upper side band is suppressed. A second modulation then translates the vestigial side-band signal back down in frequency to the final band. In contrast, the *L3* terminals employ only a single step of modulation to convert the signal directly to the assigned band as shown in Fig. 2. In general, this can be accomplished if the carrier frequency is at least half the sum of input and vestigial bandwidths. Then the lower modulation side band does not fold over the zero frequency axis to produce frequencies which fall back into the vestigial or upper side bands. Some fold-over is evident in the *L3* case shown in Fig. 2 at low frequencies of the modulator output. Single-step modulation is advantageous in that the very high frequencies encountered in the multistep process are avoided. The disadvantage of the 1-step process is that many extraneous products of modulation, which in the 2-step process can be suppressed with filters, must be reduced to tolerable levels by balances in the modulator.

The modulator, Fig. 3, is a combination of two double-balanced modulators of a form often employed for modulation of telephone signals.[4] The effect of a double-balanced varistor modulator of the type represented by either of the two in Fig. 3 is to multiply the input signal by a square-wave function having the period of the particular carrier frequency. Since the square wave contains all odd harmonic multiples of the carrier frequency, its multiplication by the input signal generates a series of double side-band output spectra, each centered about one of the harmonics of the carrier. It happens in this case that the lower side band of the third harmonic spectrum of the carrier modulator contains frequencies low enough to overlap the high frequencies of the carrier spectrum upper side band, and this overlap results in quite visible picture distortion.

It is possible, by employing a second modulator paralleling the first but driven with a frequency three times the carrier frequency, to generate a signal spectrum centered at carrier third harmonic which will cancel the corresponding output of the carrier modulator. Successful translation of the video spectrum to the *L3* carrier band in a single modulation step depends upon the maintenance of this and other modulator balances to unusually stringent requirements.

The carrier supply oscillators for the modulators and demodulators are of the Meacham bridge type[5] with quartz crystal frequency control and thermistor amplitude control. Frequency stability of two parts per 1,000,000 is required for successful carrier regeneration at the receiver. A constant temperature oven for the quartz plus the inherent stability of the bridge-type circuit is expected to provide the required frequency stability between monthly maintenance periods.

A feature of the signal transmitted over the *L3* system is a degree of modulation which exceeds the value commonly referred to as 100 per cent modulation. The resulting wave form contains a maximum ratio of information to peak carrier, important from the standpoint of optimum signal to noise performance. Fig. 4 shows progressively the reduction in peak carrier amplitude, which may be effected by subtraction of carrier component from a modulated signal. Figs. 4(B), 4(C) and 4(D) each contain the same amplitude of video modulation. Fig. 4(B) represents a video-modulated carrier signal with maximum carrier occurring at tips of synchronizing pulses, and a minimum carrier, equal to 20 per cent of maximum carrier, corresponding to picture white. Fig. 4(C) represents the same signal as Fig. 4(B) except that the 20-per-cent excess carrier has been subtracted. This is the 100-per-cent modulation case. Fig. 4(D) shows the effect of further carrier subtraction, or addition of negative carrier, to reduce to a minimum the peak amplitude of the modulated signal. The wave form of Fig. 4(D) employed for *L3* transmission requires 7½ decibels less maximum carrier power than that of Fig. 4(B) for the same transmitted information. The term excess carrier ratio has been devised to describe degrees of modulation which exceed 100 per cent. It is the ratio of peak carrier amplitude to the peak-to-peak modulation amplitude. Excess carrier ratio (ECR) for the wave forms of Figs. 4(B), 4(C) and 4(D) respectively are 1.2, 1.0 and 0.5.

Modulated signals of the forms of Fig. 4(B) or 4(C) may be detected by rectification, that is, envelope detection. However, rectification of the wave form, Fig. 4(D) produces a spurious envelope wherein video signals which exceed a particular value are inverted. It is necessary to employ homodyne detection, that is, a demodulator driven by a locally generated carrier which is synchronous in phase angle and frequency with the carrier component of the signal wave. As described later, homodyne detection also makes possible the necessary suppression of the quadrature distortion associated with vestigial side-band transmission. Quadrature distortion associated with envelope detection is tolerated in the *L1* coaxial system, but with tandem operation of terminals required in the *L3* system it would accumulate to intolerable values.

Vestigial Side-Band Considerations

A vestigial side-band signal is produced by a band-shaping filter following the modulator. In this filter the lower side band is suppressed completely except for those frequencies which are within 500 kc of TV carrier. Lower side-band frequencies within 500 kc of carrier are suppressed only partly, as also are upper side-band frequencies within 500 kc of carrier, to achieve a symmetrical response function in the vestigial side-band region.

It is convenient in a discussion of vestigial side-band transmission to consider the transmission as made up of two components, each symmetrical about carrier frequency, a real or in-phase component, and a quadrature component which is a distortion term.[6] The process is illustrated in Fig. 5. Here the response function shown in Fig. 5(A) represents idealized conditions for vestigial side-band transmission. The main side band is shown extending from carrier frequency F_c to the upper cutoff F_u. A vestige of the lower side band extends from carrier frequency to the lower cut-off F_v. Constant envelope delay is required in the entire band from F_v to F_u. In the frequency region $F_c \pm F_v$ the response characteristic is so shaped that the sum of responses at corresponding frequencies above and below carrier add to a constant value. The summing of signal components in the vestigial bands above and below carrier is accomplished by the receiver demodulator.

The response function of Fig. 5(A) may be considered to be the sum of the two response functions 5(B) and 5(C) which have even and odd symmetry respectively about the carrier F_c. Both 5(B) and 5(C) are double side-band functions. The component in Fig. 5(B) represents the normal double side-band output of the modulator supplied with video and carrier signals. The other component, Fig. 5(C), represents the output of a modulator supplied with carrier and signal

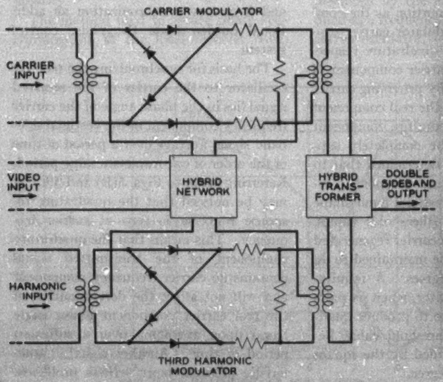
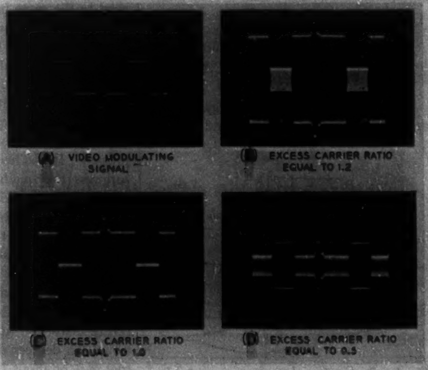


Fig. 3. Modulator for L3 terminals with third harmonic carrier product balance feature

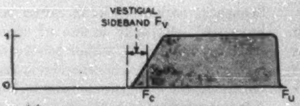

(A) VIDEO MODULATING SIGNAL

(B) EXCESS CARRIER RATIO EQUAL TO 1.2

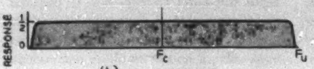
(C) EXCESS CARRIER RATIO EQUAL TO 1.0

(D) EXCESS CARRIER RATIO EQUAL TO 0.5

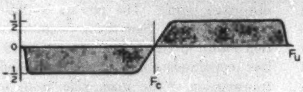
Fig. 4. Carrier waves variously modulated by a composite video signal



(a) VESTIGIAL SIDEBAND RESPONSE FUNCTION



(b) REAL COMPONENT



(c) QUADRATURE COMPONENT

Fig. 5. Response function of a vestigial side-band carrier system

voltages each shifted in phase by 90 degrees and with the amplitude attenuated as shown in the vestigial region $F_c \pm F_v$. The modulation transmitted by a circuit with the response function of Fig. 5(C) is called the quadrature component and is related to the normal modulation, depending upon the shape and extent of the vestigial side band. Fig. 6 illustrates the real and quadrature components of an idealized rectangular wave form demodulated after transmission over circuits having the response functions of Fig. 5 respectively.

The 90-degree shift of carrier frequency in the quadrature component of the vestigial side-band signal makes possible the suppression of this component. The transmitted vestigial side-band signal may be written

$$V(t)=P(t)\cos ct + Q(t)\sin ct \qquad (1)$$

where

$c=2\pi$ times carrier frequency

$P(t)$ and $Q(t)$=real and quadrature modulating functions[6] typically as represented in Fig. 6

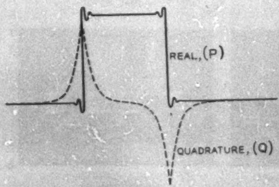


Fig. 6. Real and quadrature components of a rectangular pulse after vestigial side-band transmission

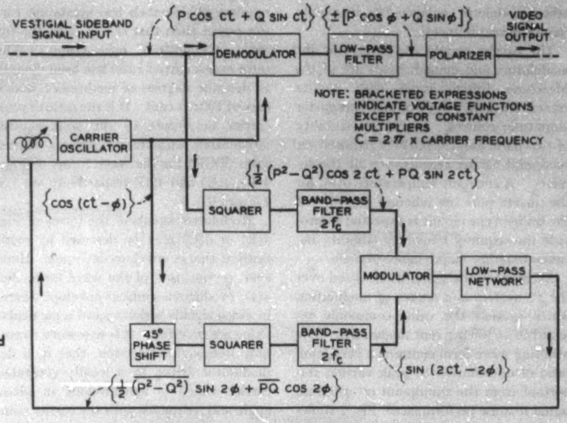


Fig. 7. Functional diagram of the homodyne demodulation process

The demodulator may be regarded as an ideal multiplier of signal and carrier supply. Let the carrier supply be denoted

$$C(t)=\cos(ct-\phi) \qquad (2)$$

where ϕ=the phase of the receiver-carrier supply relative to the carrier factor of the real component of the signal. The demodulator output is the product $V(t)$ $C(t)$. A low-pass filter rejects the output components in the band of frequencies about twice carrier frequency so that the demodulated video signal output is the lower frequency component. Thus, neglecting a factor of 1/2

$$V_0(t)=P(t)\cos\phi+Q(t)\sin\phi \qquad (3)$$

It is seen that real and quadrature video components in the demodulator output exist in the same proportion as the components of the demodulator carrier supply in phase and in quadrature respectively with the real carrier component of the vestigial signal. By providing carrier exactly in phase with the real component of the signal, the quadrature component in the output may be completely suppressed. It has been determined that to suppress the quadrature component resulting from the *L3* vestigial band shape to barely perceptible (threshold) values, the phase angle of the carrier regenerated at the receiver must be maintained to an accuracy of ±2.5 degrees. A requirement for one demodulator, when six pairs of terminals contribute to produce quadrature distortion at threshold value, becomes 2.5 degrees divided by the square root of 6, or about 1 degree.

The regeneration of carrier at the receiver is one of the principal *L3* terminal features. Here a 4.139-megacycle carrier must be provided to demodulate the overmodulated *L3* signal. The required carrier must be reconstituted from information carrier in the signal itself. It would be possible of course to transmit separately a signal from which carrier frequency could be derived, but carrier frequency is really the smallest part of the required information. It is the phase angle of the carrier of the received signal which must be duplicated closely at the demodulator and separate transmission of carrier phase angle does not seem feasible. A phase-controlled oscillator is employed for the carrier supply at the receiver, with phase control obtained from information residing in the signal itself, and frequency synchronization an additional burden upon the phase control system.

The basis for synchronizing the receiver oscillator to the carrier of the received signal lies in the phase angle of the carrier frequency component of the vestigial side-band signal average over a period of time of the order of one frame-scanning period. Referring again to Figs. 5(B) and 5(C), it may be noticed that the quadrature response function is zero at carrier frequency. This means that the quadrature component of the transmitted signal contains no carrier frequency component and will not affect the determination of the real carrier component phase angle based upon averaging over a sufficient period of time. Another signal characteristic presents more serious problems. The degree of modulation employed in *L3*, shown in Fig. 4(D), makes the average carrier polarity indeterminate. That is, the carrier polarity for a video amplitude corresponding to picture white is opposite to that corresponding to picture black or sync pulses. The polarity reverses as the composite signal changes through its half peak-to-peak value. The average polarity determined from a predominantly white picture is thus opposite to that determined from a predominantly black picture. A carrier oscillator, phase-synchronized to the average carrier phase of the signal would execute 180-degree phase reversals as picture content change from average white to average black, producing sudden video polarity reversals at the demodulator output.

This signal carrier polarity ambiguity which is momentary in character can be exchanged for one which is not time-variable by a multiplication operation. The modulated signal is squared, that is, multiplied by itself on an instantaneous basis, in a square-law circuit. Such an operation squares carrier amplitude and doubles carrier frequency and phase angle, the latter effect converting 180-degree phase reversals into 360-degree changes which are indeterminabie in the average phase detector. Under these conditions the phase-synchronized demodulator carrier supply, stably locked to the average phase of the squared signal, experiences no phase reversals with change in picture content. The ambiguity now is in the determination of incoming signal polarity. The squaring operation eliminates any basis for determining polarity so that the demodulator carrier may with equal likelihood lock to either polarity relative to the signal and thereby, at the demodulator output, produce video signal wave forms of either polarity.

The method used to secure phase synchronization of the local receiver oscillator to the received signal is described next with reference to Fig. 7. Signal from the line, together with the output from the carrier oscillator, is brought to the demodulator where the desired video output signal is obtained as the lower side band of the modulation product. This process has already been described in equations 1 to 3. In the carrier regeneration process, signal and carrier phase shifted by 45 degrees, equations 4 and 5, are each squared in square-law circuits

$$V(t)=P\cos ct+Q\sin ct \tag{4}$$

$$C(t)\underline{/45\text{ degrees}}=\pm\cos(ct-\phi-\pi/4) \tag{5}$$

Band-pass filters select from the squaring circuit output signal frequencies in the neighborhood of twice carrier frequency. From the signal squarer

$$1/2(P^2-Q^2)\cos 2\,ct+PQ\sin 2\,ct \tag{6}$$

and from the carrier squarer

$$1/2\sin(2\,ct-2\phi) \tag{7}$$

The two squared signals at twice carrier frequency are multiplied together in a product modulator. This product contains signals in two bands of frequencies, one band in the region of four times carrier frequency and the other in the video frequency band starting at zero frequency. The lower frequency component of this product is selected by a low-pass filter following the product modulator yielding

$$\frac{(P^2-Q^2)}{2}\sin 2\phi+PQ\cos 2\phi \tag{8}$$

The d-c component of the low-pass filter output is a suitable control voltage for synchronization, and is obtained in the limit as the cutoff frequency of the low-pass filter is lowered. Average values as produced by the low-pass filters are applied as a frequency control voltage to the carrier oscillator.

The first term in equation 8, $(P^2-Q^2)/2 \times \sin 2\phi$, when averaged is, for small errors in carrier phase angle, proportional to the error angle ϕ with a factor of proportionality recognized as the difference in mean-squared values of the real and quadrature modulating functions, P and Q, illustrated typically in Fig. 5. This difference is always positive when the modulating signal contains energy components within the bounds of the vestigial side band since the quadrature response function, Fig. 5(C), is attenuated relative to the real response function in this band. In the present case, with a 500-kc vestigial bandwidth and a composite video wave form for a modulating function, the amplitude of Q^2 for control purposes is negligible compared with P^2.

The second term of equation 8, $PQ \times \cos 2\phi$, contains no d-c component since Q itself contains no d-c component, and all other frequency components of Q are shifted 90 degrees in phase relative to corresponding components in P. The d-c control voltage therefore is not modified by the existence of the second term of equation 8. However, the function PQ does contain sum and difference frequencies caused by the cross products of the spectra of P and Q. These frequencies in the control voltage tend to be large compared with corresponding frequencies caused by the products P^2 and Q^2 since the trigonometric multiplier $\cos 2\phi$, equation 8, is large when the phase-angle error is small. The effect of the term $PQ\cos 2\phi$ is that of phase modulation of the receiver carrier supply, and its suppression determines the characteristics required of the low-pass filter which averages the control signal. At the penalty of sluggish synchronization and restricted oscillator pull-in range, the phase modulation can be reduced to arbitrarily small values. For our purposes a pull-in range of ±20

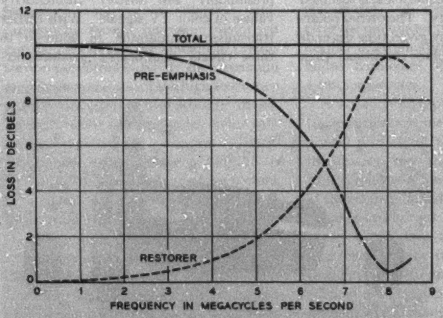


Fig. 8. High-frequency pre-emphasis characteristic

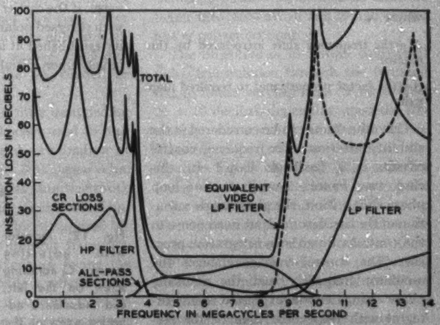


Fig. 9. Insertion loss of the transmitting terminal filters

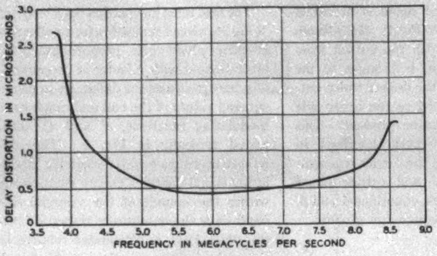


Fig. 10. Envelope delay distortion of the transmitting terminal filters

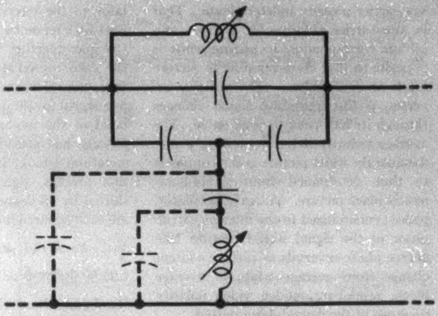


Fig. 11. Schematic of a delay equalizer section. Dotted capacitors are parasitic

cycles per second(cps) can be achieved with phase modulation less than ±0.1 degree with adequate margins.

The control voltage is applied to a tuning element in the receiver oscillator, in this case a small saturable reactor made with ferrite as a core material. This reactor is part of the series resonant quartz crystal circuit which determines the oscillator frequency, and is capable of shifting the frequency in response to the control voltage by ±20 cps, a figure chosen as safely less than the first side band components of the transmitted signal which are ±30 cycles from carrier frequency. This precaution avoids possible synchronization of the local oscillator to a signal side-band frequency, rather than to the carrier.

Sufficient gain is provided in the carrier frequency control loop just described so that the maximum frequency difference encountered between transmitting and receiving oscillators is corrected by the phase control voltage because of a steady-state phase-angle error ϕ less than 1/2 degree. The control characteristic of the saturable reactor may be expressed

$$\Delta f = A(P^2 - Q^2)\sin 2\phi \qquad (9)$$

where

Δf = the frequency shift introduced by the reactor

A = the factor proportional to required loop gain

One other factor to be considered is the stability criterion of the frequency control circuit as a feedback loop.[7] In this case two factors contribute to loop phase shift. First, the phase-angle variation of the oscillator output in response to the control voltage is an integration process. The control voltage changes the oscillator frequency and the resulting phase change can be expressed as the integral with respect to time of the frequency shift

$$\text{phase } \theta = 2\pi\int \Delta f dt \qquad (10)$$

The integration with respect to time introduces a 90-degree low-pass phase shift into the control loop at all frequencies. Also the averaging low-pass filter introduces phase shift in the same direction, so that care must be exercised to avoid instability. In this case a phase stability margin of 45 degrees is provided over a wide range of frequency by designing the low-pass filter as a series of resistance-capacitance steps of loss. These are staggered in frequency to produce a cut-off rate of 3 decibels per octave with a phase shift of 45 degrees over a wide-frequency band.

The polarity ambiguity resulting from the squaring process has been demonstrated in the derivation of the phase control voltage. In equation 5 the ± designation indicates that either polarity of carrier signal might be assumed without affecting subsequent expressions. However, in the derivation of the output voltage from the main signal demodulation, equation 3, the output signal polarity reverses if the carrier, equation 2, is assumed with reversed polarity. The carrier polarity established at any given time depends largely upon initial phase conditions when signal is applied.

Correct video polarity at the receiver output is established by a new device called a polarizer which follows the demodulator. This circuit recognized video polarity on the basis of standard features in the composite TV wave form. The particular features used in this case are the vertical blanking discontinuities expected once each 1/60 second, and the duty factor of sync pulses. These two characteristics taken together form a sufficient condition for the determination of polarity of any composite video wave form independent of picture content. The polarity, once recognized to be inverted, is corrected.

Harmonic Distortion

A significant consideration in transmission problems is the generation of distortion by the nonlinear amplitude characteristic of the transmission apparatus. In the case of video transmission, nonlinearity results in the distortion of brightness values of the transmitted picture and presumably will distort chromaticity values of color TV signals. With carrier transmission apparatus, in addition to these effects, nonlinearity produces ex-

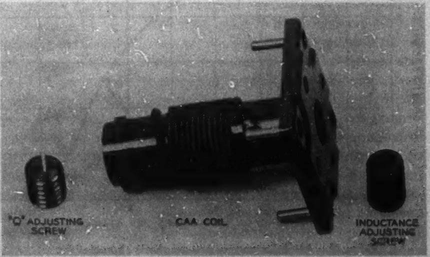


Fig. 12. Inductor with adjusting screws; Q adjusting screw on the left, and inductance adjusting screw on the right

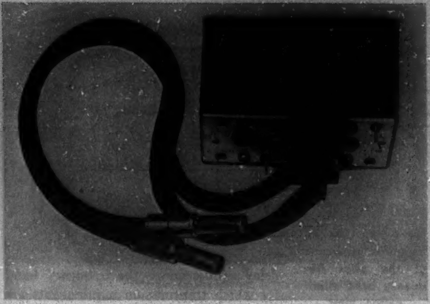


Fig. 13. Fixture for delay equalizer section adjustment

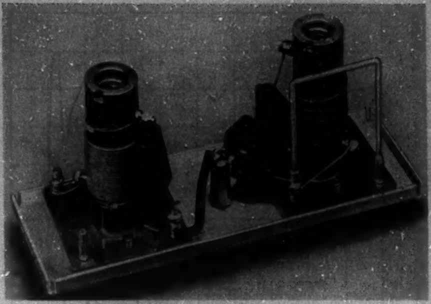


Fig. 14. Model of a typical delay equalizer section

traneous interference patterns at harmonics of the carrier frequency. In *L3*, with a carrier-modulated spectrum concentrated near 4.139 megacycles, the second harmonic distortion of line-repeating amplifiers produces a new spectrum concentrated near 8.278 megacycles, which is demodulated by the receiving terminal to the region near 4.139 megacycles. This form of distortion is considerably more disturbing in the final picture than a comparable distortion of brightness values, and constitutes a limit to transmission signal-to-noise performance.

Advantage is taken of the spectral distribution of energy of TV signals to ameliorate somewhat the effects of second harmonic distortion. A pre-emphasis network is employed in the transmitting terminal to accentuate the amplitude of high-frequency components of the signal before transmission. At the receiver, a restorer network introduces a complementary frequency characteristic to make the over-all transmission characteristic constant with frequency, Fig. 8. The restorer network, de-emphasizing the high-frequency components, likewise suppresses the second harmonic distortion signals. A limit to the amount of predistortion permitted is set by the maximum amplitudes expected of the high-frequency picture components, particularly in anticipation of a high-frequency color subcarrier in color TV systems. Tentatively, the characteristic of Fig. 8 is chosen as a compromise of these factors.

Filters and Equalizers

A low-pass filter is required before the video signal is modulated, to limit the bandwidth of the signal, eliminating possible sources of disturbing cross-products, both in the modulator and in the repeaters of the *L3* system. This filter has a cut-off at 4.3 megacycles and provides over 40-decibel discrimination to all frequencies greater than 4.8 megacycles. It consists of three *m*-derived sections and introduces about 1.3 microseconds of envelope delay distortion near its cut-off frequency. This is equalized after modulation by the delay equalizer to be described.

Following the modulator is the vestigial side-band filter which passes the upper side band, and provides 60-decibel discrimination against all frequencies in the lower side band less than 3.7 megacycles. The large discrimination is required to avoid interference with the telephone channels during transmission over the coaxial line. This filter provides a controlled loss characteristic to frequencies in the band 3.64 to 4.64 megacycles which satisfies the requirement for vestigial side-band transmission. The response function for this band is shown in Fig. 2. A flat transmission characteristic, including the effective pass band loss of the video *LP* filter, is maintained over the entire upper side band from 4.6 to 8.44 megacycles. A second low-pass filter after the modulator provides at least a 50-decibel discrimination against third and higher harmonics of the TV carrier, or 4.139 megacycles.

A 4-section high-pass filter designed by the insertion loss method[8] is used to supply 50 decibels of the discrimination at frequencies less than 3.5 megacycles. This filter has a low-pass shunt network at the input to maintain its stop-band impedance at 75 ohms. This is required because, to produce a uniform frequency characteristic, the impedance facing the modulator has to provide a reflection coefficient not exceeding 3 per cent. The remaining discrimination at frequencies less than 3.7 megacycles and the major part of the vestigial side-band shaping are provided by a group of six constant-resistance equalizer sections, as shown in Fig. 9. The low-pass filter to suppress carrier harmonics consists of two *m*-derived filter sections.

The delay distortion of the complete set of filters and loss equalizer sections is shown in Fig. 10. This includes the equivalent delay distortion of the video low-pass filter, translated in frequency for equalization after the modulator. The distortion must be equalized to a constant delay over the TV band. A delay equalizer to do this is incorporated in the filter. It was designed by a potential analogue method[9] and consists of 24 all-pass sections, each having the schematic as shown in Fig. 11. A redundant capacitor is used to avoid excessively small capacity values. The capacitors shown dotted are parasitic elements which must be compensated for by modifying the design values of the other elements. This group of sections has an insertion loss varying from 2 to 12 decibels across the pass band of the filter, caused by the dissipation of the elements. This loss, of course, must be taken into account in the design of the loss equalizer sections.

The objective is to obtain equivalent video transmission through the terminal flat to limits varying from ±.02 decibel to ±.10 decibel, depending upon the frequency characteristic of the deviation. To achieve this, close control of the dissipation loss is essential. The inductor losses which cause the major part of the delay equalizer loss characteristic vary up to ±15 per cent from nominal values. As a means for controlling inductor *Q* to ±0.5 per cent or better, a *Q*-adjusting screw is employed. The inductors are solenoids wound on molded tubes with a threaded hole through the center. A threaded magnetic dust core is used for

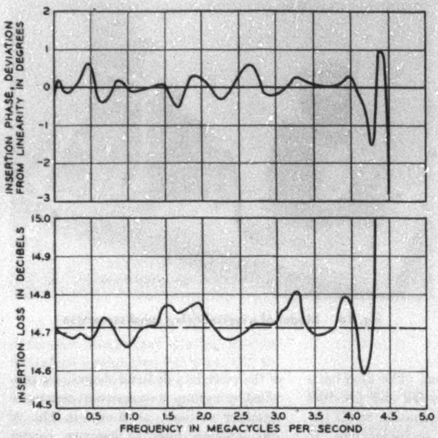


Fig. 15. Measured passband performance of a trial model of the transmitting filters and equalizer

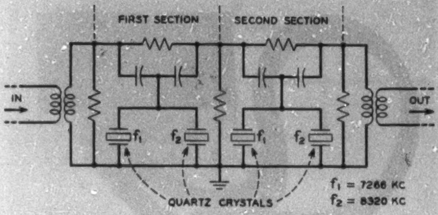


Fig. 16 (above). Simplified schematic of the pilot frequency band elimination filter

inductance adjustment. Its travel can be limited to the distance from the center to one end of the form without losing adjustment range. By introducing an additional core made of solid magnetic iron into the field of the solenoid, using the opposite end of the form, an adjustment is provided which reduces the Q in a continuous manner as the second screw is advanced into the form. The reduction in Q is caused by the losses in the iron, and normally these would cause a reduction in inductance also. However, the permeability of iron causes an increased concentration of field which tends to increase the inductance. A balance between these tendencies to decrease and to increase the inductance is obtained by controlling the geometry of the Q-adjusting core. As a result, a reduction of up to 50 per cent in Q can be obtained, accompanied by a change of less than 1 per cent in the inductance. Models of the inductor and the adjusting screws are shown in Fig. 12. This adjusting screw, in conjunction with the magnetic dust core, provides an accurate and economical means for adjusting simultaneously both inductance and dissipation in each inductor of the delay equalizer.

The flat transmission level for the upper side band and the shaped cut-off for the vestigial side band were obtained by including loss equalizer sections, assuming Q factors for the all-pass sections of about 20 per cent less than the nominal Q of the inductors. As a final step in the design, the Q factors were modified to absorb in the loss of the all-pass sections the residual loss distortion uncompensated by the loss equalizer sections. This, in effect, provided the use of 24 additional parameters for shaping the loss in the pass band, and resulted in an improved loss characteristic.

The close limits on delay distortion can be met only by close control of the adjustments on the individual all-pass sections. In order to obtain reproducible results to the order of ±0.1 degree for the phase shift and ±0.01 decibel for the insertion loss of the individual sections, a special fixture is employed to make the connection between the section and the measuring circuit. This is shown in Fig. 13. The fixture can be clamped on the network terminals quickly without soldering and provides coaxial patch cords with plugs for connection to the measuring circuit. Each section is mounted in an individual container with shielding between the inductors to reduce coupling. Each inductor is resonated with its associated capacitors, and the dissipation is adjusted by adjusting the two cores. The construction of a typical section is illustrated by Fig. 14.

An important consideration in obtaining smooth loss and delay characteristics for the delay equalizer is the reflection coefficient of each section. Poor reflection coefficients cause reflections and interactions between all-pass sections. Owing to the large phase slope of the equalizer, these tend to produce frequency characteristics with large numbers of loss and delay ripples across the frequency band for which transmission requirements are most severe. Reflection coefficients of 2 per cent or less at all frequencies in the TV band have been obtained for all delay sections by taking the following precautions:

1. Mutual coupling is limited between the two inductors in each section by use of a shield in the section container. As little as 0.1 per cent coupling coefficient can cause a reflection coefficient of 1 per cent in certain sections at the frequency of a 180-degree phase shift.
2. The Q factors of both inductors are adjusted to be equal in each section.
3. The values of the capacitors are modified to compensate for the presence of the parasitic capacitances associated with the shunt arm, Fig. 11.

The measured insertion loss characteristic and the phase shift deviation from linear phase slope for one model of the transmitting filter and delay equalizer are shown on Fig. 15. This has been reduced to video frequency to show the detailed residual distortion which results from the addition of the vestigial lower and upper side band. The delay equalization is maintained for about 200 kc above the loss cut-off to provide for at least a 30-decibel insertion loss at frequencies where the delay distortion becomes large. Without this precaution the transient response is characterized by a severe cut-off ring distortion which is a slowly damped oscillation at cut-off frequency generated by high-frequency signal components.

Variations in transmission response in the region near TV carrier cause noticeable smear distortion on the received images. These variations can be introduced by small changes in element values of the filter or equalizer sections following initial adjustment, or by other changes in both transmitter and receiver. As a means of compensating for such variations, two adjustable loss equalizers have been provided, each with adjustable maximum loss at carrier frequency. One has a half-loss point at 300 kc and the other at 80 kc from the carrier. These

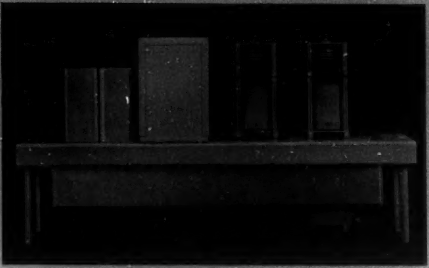


Fig. 17. Model of the carrier suppression filter

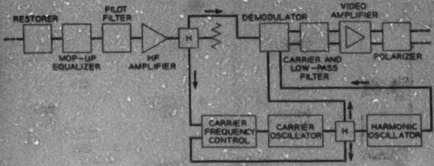


Fig. 19. L3 receiving terminal block diagram

are adjusted for minimum smear at periodic intervals.

A fixed equalizer is provided in each terminal to compensate for loss and delay distortion other than that in the filters described. For convenience, this equalization is done at modulated frequencies. It is expected that additional means for periodic re-equalization of the terminal gain and phase characteristics will be necessary to achieve 4,000-mile, tandem terminal transmission quality objectives.

Pilot Frequency Filters

As described elsewhere,[1] the *L3* system has six pilot frequencies for regulating automatically the transmission characteristic of the line. Two of these, 7.266 megacycles and 8.320 megacycles, are in the TV band. At the transmitting terminal a pilot elimination filter for these two frequencies is required to prevent energy in the TV signal near these frequencies from disturbing the pilot levels on the system. At the receiving terminals, these pilot frequencies must be removed from the TV band to avoid interfering effects in the output TV signal. A discrimination is required of 50 decibels for frequencies within 20 cycles of the two pilot frequencies. Also, at the receiver a carrier elimination filter is required to suppress the residual carrier leak from the demodulator. This demands a 30-decibel suppression to frequencies within 20 cycles of 4.139 megacycles.

To avoid removing excessively broad bands of frequencies from the television band, and thereby generating visible distortion in the transmitted picture, narrow bandwidth crystal filters are used for both of these applications. The pilot elimination filter has its 3-decibel loss points at about 1,000 cycles on either side of the pilots. The carrier suppression filter has its 3-decibel loss points at 150 cycles on either side of the carrier. In the past, spurious or secondary responses in the crystal units could not be limited to sufficiently low values to permit available crystals to be used in broad-band circuits without elaborate means for suppressing the unwanted responses. In the *L1* system, for example, a balancing circuit using hybrid coils was employed for this function. Techniques for reducing unwanted responses in the crystal units by special contouring of the blanks and by precise optimum area plating have been developed recently.[10] The use of these methods has resulted in crystal units in which the unwanted responses are reduced to the extent that relatively simple filters can be used for these applications. A simplified schematic of the pilot elimination filter is shown in Fig. 16. Small ovens are provided to maintain close temperature control for the quartz crystal elements in order to stabilize the resonant frequencies. Mechanical arrangements are illustrated in Fig. 17.

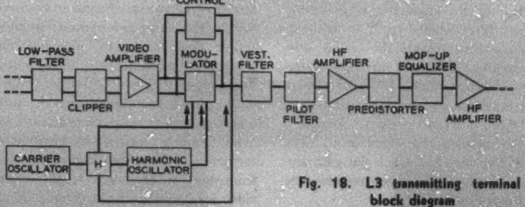


Fig. 18. L3 transmitting terminal block diagram

Power Equipment

Primary power for the TV terminals is 60 cycles a-c. It is derived from the *L3* motor-alternator power equipment at main repeater stations or, alternatively, from commercial 60-cycle sources. When commercial power is used directly, a provision is made for switch-over to emergency power in the event of power failure. The emergency power supply comprises a 130-volt battery, an inverter, and automatic switch-over equipment. If during service the commercial power source fails, a switch-over to emergency power is made automatically in a period under 1/10 second.

D-c and heater power provisions involve two departures from previous practice. First, the rectifiers which supply anode and other d-c power requirements are nonregulated metallic (selenium), rectifiers. These are supplied from magnetic a-c line voltage regulators to obtain suppression of line voltage variations. The combination of the magnetic a-c regulator and the metallic rectifier provides adequate suppression of hum and line voltage variations with, it is expected, greater reliability than alternative vacuum-tube regulated rectifiers. Secondly, the magnetic a-c line voltage regulator, together with an improved heater transformer design, makes possible the operation of the vacuum tubes at reduced heater voltage to obtain increased thermionic life expectancy. This advantage can be taken only if close control of heater transformer output voltage is possible. In this case, variations in output voltage expected to occur, owing to temperature variations of the transformer windings, have been substantially eliminated by incorporating thermistor temperature compensation in the transformer primary circuit.

T3 Transmitting Terminal

A block diagram of the components of the transmitting terminal is shown in Fig. 18. The composite video signal is received from the Bell System TV operation center over a 124-ohm balanced

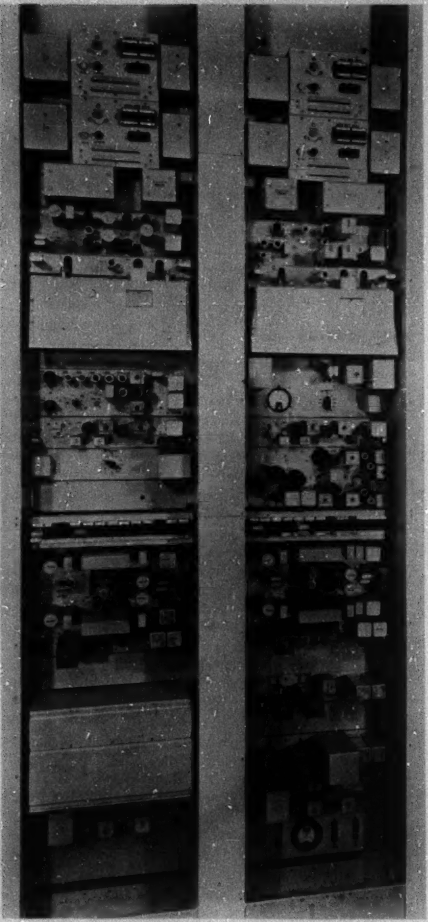


Fig. 20. Model of transmitting and receiving terminal equipments

cable, and the modulated output signal is delivered to the *L3* line facilities over a 75-ohm coaxial cable.

An input low-pass filter and a video clipper circuit place ceilings on the maximum transmitted signal bandwidth and signal amplitude respectively. The clipper circuit is required to protect the 600 telephone channels from inadvertent overloads caused by excessive TV signals. Normal amplitude signals are not affected by the clipper.

Following are a video amplifier and the modulator, together with the required carrier supplies, and the carrier level control. The latter device accurately regulates the peak carrier magnitude in the output signal and thereby preserves the desirable maximum modulation.

Following the modulator is the vestigial band filter which, together with its delay equalizer, shapes the double sideband modulator output into the vestigial transmission band. The flat loss of these networks requires that amplification be provided in the carrier frequency band to increase the signal level. The first of two high-frequency flat-gain amplifiers in the transmitter restores the signal amplitude.

The pilot band elimination filter is next provided to remove TV signal energy and other possible interference from particular frequencies allocated to the *L3* line pilot signals.

At this point occur the predistorter network for pre-emphasizing the high-frequency signal components, and a mop-up equalizer. The mop-up equalizer is to provide means for periodic correcting of the transmission characteristic. The high-frequency amplifiers, particularly, change their transmission characteristic as the vacuum tubes age. An additional high-frequency amplifier is provided to recover the loss of these networks and deliver a proper signal level to the line equipment.

R3 Receiving Terminal

The receiver demodulates the signal as transmitted over the *L3* line facilities to recover the video signal for transmission over local video circuits. As has been discussed, the demodulation is a homodyne process utilizing a local carrier regenerated from information contained in the transmitted signal.

A block diagram of receiver components is shown in Fig. 19. The first components are a group of networks; the restorer to compensate for transmitter predistortion, the fixed and variable mop-up equalizers to compensate for deviations in the receiver frequency characteristic, and a pilot elimination filter to remove *L3* pilot frequencies from the signal. A high-frequency amplifier provides amplification to compensate for the network losses.

The amplified signal is split into two branches by a hybrid transformer. Part of the signal serves as a carrier to the demodulator for detection to video frequencies, while the remainder is employed in the carrier regeneration apparatus. The carrier supply oscillator output likewise is split between the same two circuits, the portion supplied to the demodulator providing the carrier energy for demodulation, and the part sent to the carrier regeneration circuits providing means for comparison with the input signal. The carrier regeneration equipment comprises mainly the phase comparison circuit together with means for changing electronically the frequency of the carrier supply oscillator.

The phase-synchronized carrier supply, together with the associated third harmonic supply, provides demodulating carriers which translate the carrier spectrum down in frequency to the original video spectrum. The demodulator output also contains an unwanted upper side band in the 8-to-12 megacycle region. This signal and the residual carrier leak from the demodulator are removed from the detected signal by a low-pass filter and the quartz crystal narrow-band rejection filter tuned at carrier frequency. Finally, the signal is amplified at video frequencies to provide moderately high video frequency transmission levels and is transmitted through the reversing relays of the polarizer to the receiver output.

Fig. 20 is a photograph of two equipment bays, each 11 feet high, which comprise at the left the *T3* transmitting terminal, and at the right the *R3* receiving terminal.

References

1. The L3 Coaxial System, **C. H. Elmendorf, R. D. Ehrbar, R. H. Klie, A. J. Grossman.** *AIEE Transactions*, vol. 72, pt. I, Sept. 1953, pp. 395–413.

2. L3 Coaxial System—Amplifiers, **L. H. Morris, G. H. Lovell, F. R. Dickinson.** *AIEE Transactions*, vol. 72, pt. I, Nov. 1953, pp. 505–17.

3. Television Terminals for Coaxial Systems, **L. W. Morrison, Jr.** *AIEE Transactions*, vol. 68, pt. II, 1949, pp. 1193–99.

4. Copper Oxide Modulators in Carrier Telephone Systems, **R. S. Caruthers.** *Bell System Technical Journal*, New York, N. Y., vol. 18, April 1939, pp. 315–37.

5. Bridge-Stabilized Oscillator, **L. A. Meacham.** *Bell System Technical Journal*, New York, N. Y., vol. 17, Oct. 1938, pp. 574–91.

6. Certain Topics in Telegraph Transmission Theory, **H. Nyquist.** *AIEE Transactions*, vol. 47, pt. II, Feb. 1928, pp. 617–44.

7. Network Analysis and Feedback Amplifier Design (book), **H. W. Bode.** D. Van Nostrand Company, Inc., New York, N. Y., 1945.

8. Synthesis of Reactance 4-Poles, **S. Darlington.** *Journal of Mathematics and Physics*, Cambridge, Mass., vol. 18, Sept. 1939, pp. 257–353.

9. The Potential Analogue Method of Network Synthesis, **S. Darlington.** *Bell System Technical Journal*, New York, N. Y., vol. 30, April 1951, pp. 315–65.

10. High Frequency Crystal Units for Use in Selective Networks and Their Proposed Application in Filters Suitable for Mobile Radio Channel Selection, **L. J. La Brie, D. F. Ciccolella.** *Transactions*, Institute of Radio Engineers, New York, N. Y., PGVC-3, June 1953, p. 118.

No Discussion

A Universal Meter for Measuring Voltages at High Impedances, Micromicroamperes, and Insulation Resistance

W. R. CLARK
MEMBER AIEE

R. E. WATSON
ASSOCIATE MEMBER AIEE

G. C. MERGNER
ASSOCIATE MEMBER AIEE

Synopsis: This paper describes a multipurpose instrument which can measure direct voltages in high-impedance circuits, small direct currents, and high resistances with good accuracy. The basic detector is a chopper-type stabilized feedback amplifier which is entirely a-c powered.

General Considerations

IT IS desirable to measure a voltage by a null method to avoid errors due to drawing current from the voltage source. It is also desirable to measure a current by a null method, to avoid any error due to a change in current caused while inserting the resistance of the meter in the current circuit.

By using a chopper-type feedback amplifier, a null-type measurement of voltage and current can be made with good accuracy. Zero stability is achieved by the chopper type of measurement, and gain stability is achieved by having sufficient gain in the feedback amplifier.[1]

To make resistance measurements, the only additional requirement is a stabilized direct voltage of known value to connect in series with the unknown resistance in order to supply the current for measurement. The instrument to be described has all of these features.

Versatility of the Instrument

The instrument, Fig. 1, features an easily read meter, with two scales approximately 5¹/₂ inches in length. One scale is for voltage and current, and the other for resistance. Two controls are provided, one to select the range of measurement and the other to select the type of measurement: voltage, current, or resistance. These controls make available ranges which have the full-scale values listed in Table I. The type of measurement employed eliminates the need for a zero adjusting control.

Two binding posts are provided for connecting the instrument for measurements. One of these does not require shielding, but for extremely low-level measurements, shielding of the other binding post is necessary. A receptable for a shielded plug is also provided for convenience in making connections to the sensitive terminal. Terminals are provided (*E* in Fig. 1) so that an external battery can be inserted in the circuit in order to measure resistance at any desired voltage level up to 500 volts.

To protect the user, the push button switch "*M*Ω" in its normal position applies only 10 volts to the input circuit for measuring resistances. It has to be pushed to make a measurement using 100 volts. The instrument may be used with a recorder for all ranges.

Description of the Basic Circuit

The basic circuit is shown in block diagram in Fig. 2 and schematically in Fig. 3. The input signal is converted from a direct voltage to a 60-cycle voltage by a mechanical synchronous chopper. Since it is desirable to have a high-input

Table I. Ranges of the Universal Meter

Volts	Microamperes	Megohms At 10 Volts	At 100 Volts
0.5	5.0	2×10^2	2×10^3
5.0	5×10^{-1}	2×10^3	2×10^4
50.0	5×10^{-2}	2×10^4	2×10^5
500.0	5×10^{-3}	2×10^5	2×10^6
	5×10^{-4}	2×10^6	2×10^7
	5×10^{-5}	2×10^7	2×10^8

Paper 53-298, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 18, 1953; made available for printing May 5, 1953.

W. R. Clark, R. E. Watson, and G. C. Mergner are with the Leeds and Northrup Company, Philadelphia, Pa.

The assistance of A. J. Williams, Jr., and W. G. Amey in reviewing the paper is greatly appreciated.

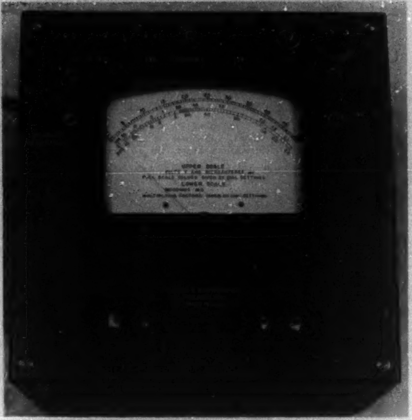


Fig. 1. The universal meter

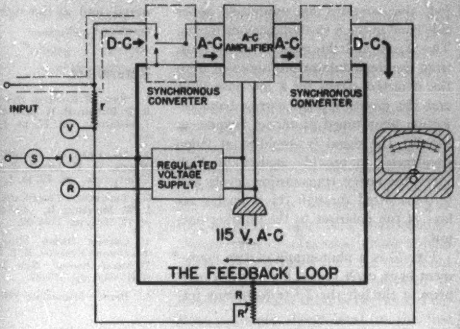


Fig. 2. Block diagram of universal meter

impedance, the input tube of the amplifier is used as a cathode follower. Following this are three stages of voltage amplification.

The resulting amplification makes the power level at the input stage so low that a d-c supply must be used for the heaters of the tube to avoid a-c pick-up from this source. It is convenient to use a d-c supply on the heater of the second tube, since this procedure serves a dual purpose: it heats the tube cathode, and it helps to filter the d-c supply for the first tube.

All tubes and their associated components thus far mentioned are contained in a shielded compartment, with external connections made through lead-through capacitors. This minimizes the effects of the radio-frequency components of line transients.

The second or output synchronous converter is electronic, and uses a duotriode tube. The grids of both tubes are connected in parallel to the output of the last voltage amplifier stage. The plate of one triode is supplied with a 60-cycle voltage 180 degrees out-of-phase with the plate supply of the other triode. The converter triode having its plate voltage in phase with its grid voltage will conduct greater current than the other triode, which at this time will have its plate voltage 180-degrees out-of-phase with its grid voltage. The voltages produced by the corresponding plate currents flowing through the cathode resistors of the tubes are added to produce the voltage output of the electronic converter. This voltage, after being filtered to minimize the ripple, is fed back into the input circuit to stabilize the gain of the instrument.

If the voltage is fed back in series with the detector and the input to be measured, the instrument makes a null-type measurement of voltage. Referring to Fig. 2, this occurs when switch *S* is in the *V* position and *r* is open circuited. In this case, the voltage range is determined by the voltage drop across *R* produced by the current flowing in the meter.

If the voltage is fed back through a resistor and this combination is in parallel with the detector and the input to be measured, the instrument makes a null-type current measurement. This occurs when the switch *S* of Fig. 2 is in the *I* position. In this case the current range is determined by the voltage drop across *R*, divided by the value of *r*.

In order to make a resistance measurement, it is only necessary to apply a known regulated voltage to the unknown resistor, and measure the current through the resistor by the null-current method previously described. Such a regulated supply is built into this instrument. To measure resistance it is only necessary to turn switch *S* to the *R* position.

Sufficient stability is obtained for the regulated supply by using two successive stages of regulation, each stage employing a glow discharge type of regulator tube. The first stage used a type-OA2 regulator tube and the second stage uses a type-5651 voltage reference tube. Since only 1 milliampere is required, excellent regulation is obtained by this method.

Four voltage ranges are provided. The two low ranges are obtained by selecting the proper value of *R*, and hence, a direct measurement is made of the unknown voltage. For the higher ranges of voltage, a high-resistance voltage divider is incorporated in the instrument and used to attenuate the voltage to a value which is then measured by the low range method.

The six current ranges and six resistance ranges are obtained by changing the values of *r* and *R*; see Fig. 3.

The push-button switch which changes the regulated voltage applied to the unknown resistance from 10 volts to 100 volts provides a method of quickly checking the voltage coefficient of resistance, as well as supplying the higher resistance range. The instrument is calibrated for the high-resistance range using the 100-volt supply.

The type of circuit used makes the equivalent input resistances for the voltage ranges quite high, the equivalent input resistances for the current ranges fairly low, and greatly reduces the effect of input capacity in resistance measurements. For operating convenience, a 10,000-megohm resistor is connected across the input for all voltage ranges.

Terminals are provided so that a null-type voltage recorder with a range of 20 millivolts can be connected across part of the *R* resistor and used for all ranges when it is desired to record measurements. A filter is incorporated in the input circuit of the amplifier to minimize the effects of spurious a-c pick-up in the input leads.

In order to make measurements at high impedances, good insulation must be provided. Specially designed switches and terminals are employed using for insulation an acrylo-nitrile copolymer of styrene. These switches are leaf type,

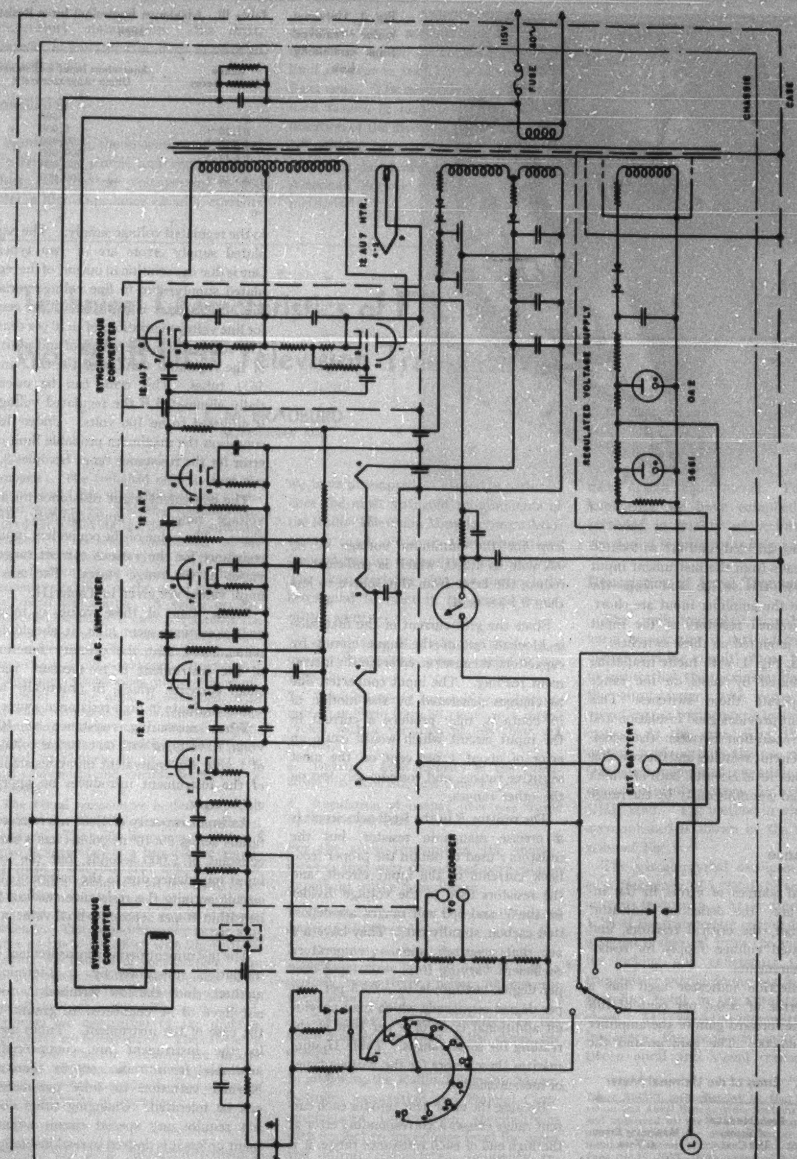

115V
60~
FUSE
CHASSIS
CASE
12 AU 7 HTR.
SYNCHRONOUS CONVERTER
12 AU 7
A.C. AMPLIFIER
12 AX 7
12 AX 7
SYNCHRONOUS CONVERTER
REGULATED VOLTAGE SUPPLY
OA 2
5651
EXT. BATTERY
TO RECORDER
L
H
G

Fig. 3. Schematic diagram of universal meter

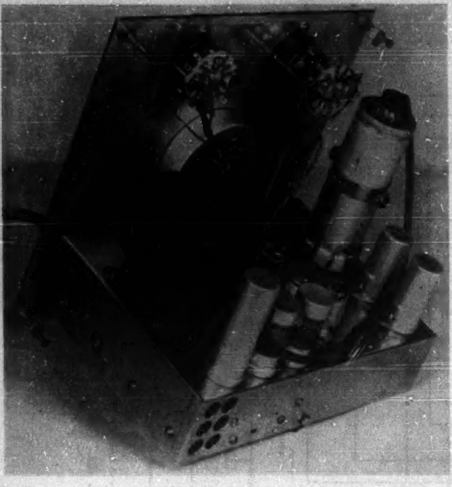


Fig. 4. Universal meter removed from containing box

isolated and shielded, and are so located that the leads from the instrument input terminals to the switch and from the switches to the amplifier input are short. The high-valued resistors of the input circuit are mounted on these switches.

Plungers, Fig. 4, with lucite insulating caps, activated by cams on the range switch, operate these switches. This arrangement provides good insulation and sufficient separation between the sensitive input-circuit switches and the switches of the higher level circuits, both of which are operated simultaneously by the range control.

Performance

Principal sources of errors in the instrument are: the deflection indicator, the amplifier, the critical resistors, and the regulated voltage supply for resistance measurements.

The deflection indicator used has a limit of error of ±0.5 per cent of full scale. The forward gain of the amplifier is about 65,000. The gain around the loop for the minimum voltage at R, 0.5 volt, is 1,200, which is sufficient to reduce the error from this source to less than 0.1 per cent.

Table II. Errors of the Universal Meter

Type of Measurement	Probable Limit of Error, Per Cent	Maximum Error at Zero
Voltage	1.5 of full scale	0.005 volt, on 0.5 and 5.0 volt ranges 0.5 volt, on 50 and 500 volt ranges
Current	3.5 of full scale	10^{-12} ampere
Resistance	5.0 of reading	

Since the grid current of the first tube is blocked out of the input circuit by capacitors, it causes no errors in the instrument reading. The input converter, due to charges generated by the motion of its contacts, may produce a current in the input circuit which would cause an error of about 1 per cent on the most sensitive ranges, and considerably less on the other ranges.

The resistor R in the feedback circuit is a precise manganin resistor, but the resistors r used to obtain the proper feedback currents in the input circuit, and the resistors used in the voltage divider for the 50 and 500 volt ranges are deposited carbon, spirally cut. They have a 1 per cent accuracy, plus a temperature coefficient varying from −0.04 per cent per degree centigrade to −0.15 per cent per degree centigrade which may provide an additional error of 3 per cent of the reading for some ranges. Table II summarizes these errors for the various types of measurements.

Because the error near zero for each current range causes a corresponding error at the high end of each resistance range, it is recommended that, for resistance measurements, a range be selected which will give a scale reading of less than 20. The resistance error includes those errors associated with the current ranges, plus errors due to the regulated voltage supply. The regulated supply errors are of two types. One is due to variation in output of the regulated supply, due to line voltage variation. This error is less than 0.1 per cent for line voltage variations of ±10 per cent. The other error is due to lack of uniformity of the regulated voltage of the *OA2* and *5651* tubes. This error can be essentially eliminated if the regulated voltage is adjusted to be 100 volts. Under this condition the maximum probable limit of error for the resistance range becomes 3.6 per cent.

Table III. Maximum Equivalent Input Resistance

Range Microamperes	Equivalent Input Resistance, Ohms (Approximately)
5	500
5×10^{-1}	5,000
5×10^{-2}	50,000
5×10^{-3}	500,000
5×10^{-4}	4,900,000
5×10^{-5}	45,000,000

The equivalent input resistance for all voltage ranges is approximately 10^{10} ohms. The value of the equivalent input resistance for the various current ranges varies as the range varies. The maximum values are given in Table III.

While some of these values of input resistance may seem high, it should be remembered that the primary function of the instrument is to measure very small currents, which, in practically all cases, originate in high-resistance sources.

When measuring resistance at 100 volts, 10 volts, or with an external voltage of 1 volt, the equivalent input resistance of the instrument introduces no appreciable error.

An input capacity of 1,000 micromicrofarads across 2×10^6 megohms has a time constant of 2,000 seconds, but the low input impedance due to the nature of the circuit permits the resistance reading to be within 10 per cent of its final value in 8 seconds.

The instrument permits measurement of insulation resistance by a 3-terminal method, since the low terminal L does not have to be connected to ground or the case of the instrument. Tubes used in the instrument are commercially available from radio supply dealers. Normal variation in tube parameters can be tolerated. Changing tubes does not require any special circuit adjustment unless it is desired to read insulation resistance to better than the accuracy stated above.

The instrument has a warm-up time of approximately 1 minute, but there will be some slight drift thereafter due to temperature stabilization. This drift, however, is less than the limit of error of the instrument.

Conclusion

Summarizing, the instrument described is a relatively simple and versatile unit, which can measure voltages up to 500 volts in high-impedance circuits, currents as low as 10^{-13} ampere, and insulation resistances up to 10^8 megohms, with a limit of error in each case not exceeding 5 per cent. The method of measurement used results in negligible errors due to insertion of the instrument in the circuits. It permits measurement of insulation resistance in accordance with approved American Society of Testing Materials methods.[2]

References

1. D-C Amplifier Stabilized for Zero and Gain, A. J. Williams, Jr., R. E. Tarpley, W. R. Clark. *AIEE Transactions*, vol. 67, pt. I, 1948, pp. 47–57.

2. Proposed Revised Tentative Methods of Test for Electrical Resistance of Insulating Materials. *Publication No. D-257-52T*, American Society for Testing Materials, Philadelphia, Pa.

◆

No Discussion

Technical Characteristics of FTL Type No. 20-B UHF Television Transmitter

E. M. BRADBURD
NONMEMBER AIEE

Synopsis: The technical objectives of the design of a 1-kw ultrahigh-frequency television transmitter and the methods employed for achieving the desired results are outlined. The aural transmitter is described and details are given for the design of the frequency-lock system, which insures frequency stability of the aural relative to the visual transmitter. A single crystal-controlled oscillator is used to reference the frequency-modulation aural oscillator, and to provide the necessary frequency offset between aural and visual carrier frequencies. The frequency-multiplier and power-amplifier chain in the aural is described, with particular emphasis on the reasons for the engineering choices made. Power amplification and frequency multiplication are performed simultaneously in order that the most economical and reliable design can be obtained.

The visual transmitter is described, and details of the modulation method employed are given. The radio-frequency modulated signal is generated in a broad-band amplifier stage employing grid modulation. A double-tuned output circuit is employed to realize the necessary broad-band characteristic, and to obtain good amplifier circuit efficiency. The radio-frequency power amplifier design is described, with illustrations showing the method of construction. The vestigial sideband filter and diplexer units are analyzed and the method of construction of the necessary cavities is described. Formulas for calculation of the diplexer parameters are briefly given. The cabinetry cooling and control circuitry are briefly described to illustrate the methods employed to insure ruggedness, ease of access and automaticity of operation.

THE FTL 20-B ultrahigh-frequency television transmitter is designed to furnish a nominal synchronous peak-power output of 1 kw in the ultrahigh-frequency bands from 470 to 890 megacycles. In the design of this equipment, we have attempted to exceed in performance the most stringent requirements of the Radio-Television Manufacturers Association and Federal Communications Commission for transmitters of this type. The characteristics that we particularly attempted to meet in the visual transmitter were:

1. Carrier-frequency stability better than one part per million over a period of one week;
2. Video-frequency response (visual transmitter) uniform from 30 cycles to 4 megacycles, with a variation of less than ±1 decibel;
3. Deviation from linearity of the modulation characteristic less than 10 per cent;
4. Regulation of output over one frame less than 4 per cent.

For the aural transmitter, the characteristics that were considered particularly important in the design were:

1. Distortion less than 1 per cent, from 50 cycles to 15,000 cycles.
2. Frequency-modulation and amplitude-modulation noise level 60 decibels and 55 decibels respectively, below 100-per-cent modulation.
3. Frequency variation relative to the visual carrier less than ±100 cycles per second.

In addition, the Radio-Television Manufacturers Association and Federal Communications Commission specifications with respect to impedance level for input and output connections, frequency response of the aural transmitter, linearity for the visual transmitter, constancy of black level, etc., were all carefully considered as design targets for this equipment.

We considered it quite important that the equipment utilized to the greatest extent possible identical tube types and circuits in the aural and visual transmitters; that provisions for rapid servicing be included and that the design be capable of extension for use with higher power amplifiers. In the following description of the equipment we will attempt to show how we have met these equipment design requirements. To our best knowledge we have complied with or exceeded in quality of performance all the design requirements listed above.

Description of Aural Transmitter

The radio-frequency signal for the aural transmitter is derived from a frequency locking circuit, whose function is to insure that the maximum deviation of the aural-carrier, with respect to the visual-carrier frequency, is less than ±100 cycles. At the same time, the locking system used permits the use of a single frequency-modulated modulator and frequency-control system for the entire UHF band. The method in which this is accomplished is shown in the block diagram of Fig. 1.

The visual crystal oscillator operates at a frequency of 1/96 of the transmitter output frequency. At a frequency of 1/24 of the transmitter output frequency, a signal is derived from the visual chain for locking and generating the aural-transmitter carrier frequency for a corresponding point in the aural-transmitter chain. At 1/24 of the transmitter output frequency, the frequency separation between aural and visual transmitters is

---

Paper 53-299, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted April 6, 1953; made available for printing May 13, 1953.

E. M. Bradburd is with the Federal Telecommunication Laboratories, Lodi, N. J.

The author wishes to acknowledge the support and effort of M. Silver, R. Troell, R. Alter, and many others, without whom this work would not have been possible.

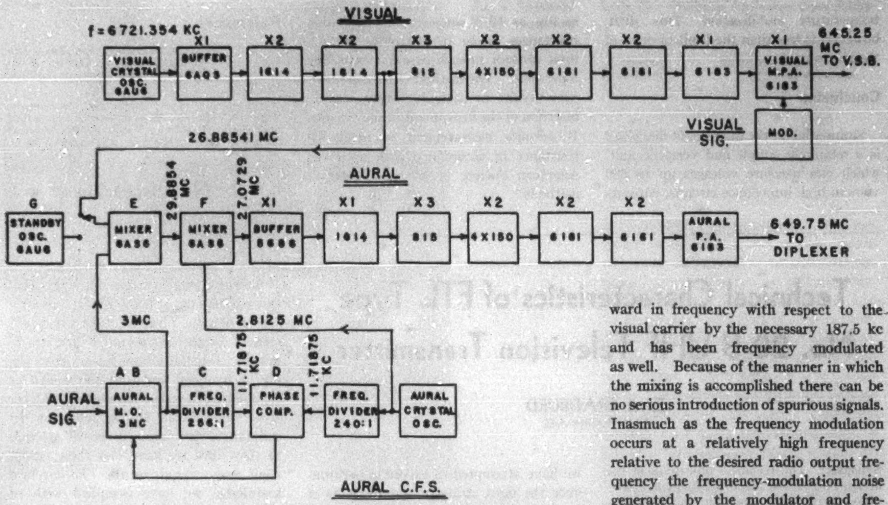

VISUAL
f=6721.354 KC
X1
X2
X2
X3
X2
X2
X2
X1
X1
VISUAL CRYSTAL OSC. 6AU6
BUFFER 6AQ5
1614
1614
815
4X150
6161
6161
6163
VISUAL M.P.A. 6163
645.25 MC TO V.S.B.
VISUAL SIG.
MOD.
26.88541 MC
AURAL
29.8854 MC
27.0729 MC
G
STANDBY OSC. 6AU6
E
MIXER 6AS6
F
MIXER 6AS6
X1
BUFFER 5686
X1
1614
X3
815
X2
4X150
X2
6161
X2
6161
AURAL P.A. 6163
649.75 MC TO DIPLEXER
3MC
2.8125 MC
11.71875 KC
11.71875 KC
A B
C
D
AURAL SIG.
AURAL M.O. 3MC
FREQ. DIVIDER 256:1
PHASE COMP.
FREQ. DIVIDER 240:1
AURAL CRYSTAL OSC.
AURAL C.F.S.

Fig. 1. Block diagram FTL 20-B 1-kw ultrahigh-frequency television transmitter

187.5 kilocycles. Therefore, the object of this locking circuit will be to add to the signal derived from the visual transmitter a frequency of 187.5 kc, which is suitable frequency modulated. Referring to Fig. 1, we see that the frequency-modulated oscillator *A* operates at a frequency of 3 megacycles. This oscillator, which is modulated by a Miller-effect modulator *B*, feeds its output to a counter circuit C, employing stable phantastron-type counters. The counter divides the input frequency by 1/256, and yields an output frequency of 11.719 kc.

The output of the frequency counter *C* is, in turn, fed to a phase-comparator circuit *D*. A reference crystal oscillator, operating at 2.8125 megacycles, is similarly fed to a counter-type frequency divider whose ratio is 1-to-240, again yielding 11.719 kc, and this signal likewise is fed to the phase-comparator circuit *D*. The error signal from the phase comparator, after suitable filtering and mixing with the desired audio modulation, is fed to the Miller-effect modulator for modulation and control of the aural oscillator. In this way, the aural oscillator, operating at a frequency of 3 megacycles, is held accurately in lock to the aural reference crystal, and is frequency modulated in accordance with the desired audio signal. The maximum deviation of the frequency-modulation signal, at this point in the chain, is approximately ±1.5 kc. This 3-megacycle frequency-modulation signal, is fed to a mixer circuit *E*, where it is heterodyned with the signal derived from the visual-transmitter crystal oscillator to yield the upper-sideband product. A suitable filter chain is used to select the desired upper side-band. It is important to note here that, because the ratio of the frequencies of the signals being mixed is always greater than 6, (hence the lowest-order undesirable-mixing product is sixth-harmonic) freedom from spurious signals and so-called cross-over points is obtained.

Thus, if the visual carrier frequency were 645.25 megacycles, the signal delivered to the sum-mixer *E* from the visual chain would be 26.8854 megacycles, and the output of the sum mixer would be 29.8854 megacycles. The output of the sum mixer is then fed to a difference mixer *F*, where the 2.8125 megacycles generated by the aural reference crystal is heterodyned with the 29.8854 megacycle-signal to yield a frequency of 27.0729 megacycles. We have achieved, therefore, a signal which has been shifted upward in frequency with respect to the visual carrier by the necessary 187.5 kc and has been frequency modulated as well. Because of the manner in which the mixing is accomplished there can be no serious introduction of spurious signals. Inasmuch as the frequency modulation occurs at a relatively high frequency relative to the desired radio output frequency the frequency-modulation noise generated by the modulator and frequency-control process is relatively insignificant. The precision of the lock between the aural and visual transmitters is a function of the stability of the aural reference oscillator alone. Inasmuch as a crystal having an accuracy of 10 parts per million is readily obtained, the precision of the lock between the aural and visual transmitters is easily made better than ±100 cycles.

In the event of a failure of the visual crystal oscillator, the standby oscillator *G* is provided so that operation for the aural transmitter is possible, with somewhat reduced frequency stability. The aural-frequency modulator and locking system is contained on a single chassis, with all tubes accessible at the front of the transmitter. Test points and meters have been provided for rapid servicing of the counters and lock chain.

The output of the aural-frequency locking unit is fed to a single stage of amplification employing a type *1614* tube. The balanced output of the *1614* amplifier is used to drive a type *815*

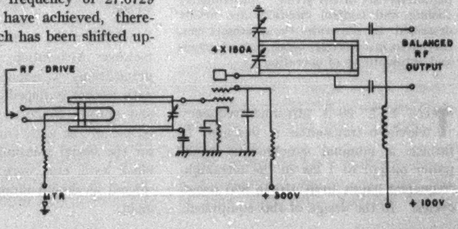

RF DRIVE
4X150A
BALANCED RF OUTPUT
MTR
+300V
+100V

Fig. 2. Balanced transmission line radio-frequency circuit 4X150 doubler

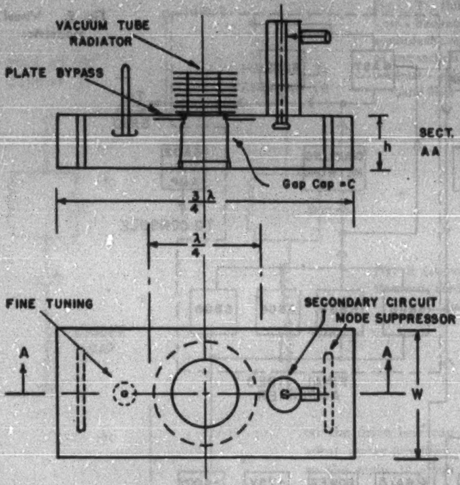


Fig. 3. Matching conditions at output of type 6183 tetrode amplifier

The resonant frequency f for the cavity is given by

$$\frac{1}{2\pi fC}=754\frac{fh}{c}\times\tan\frac{wf\pi}{c}$$

c = velocity of light
C = gap capacitance

tube in a tripler circuit, yielding a frequency equal to 1/8 of the desired aural transmitter output frequency. The radio-frequency power level at this point in the system is approximately 1 watt. This power is used to drive the first frequency doubler, employing a *4X150A* tube in a pseudo-balanced circuit. The power output of this stage is 25 watts. It is important to note here that this tube is a critical one in the system, because of its use to yield a large power gain and also because of the fact that it is the last grid driven radio-frequency stage in the transmitter. In effect, all the remaining stages (all of the grounded-grid type) are in series with the output circuit of the *4X150A* doubler. Because of this, great care was taken to stabilize the gain of this doubler by large amounts of d-c cathode degeneration and a screen supply of poor regulation. All mechanical parts are ruggedly constructed to prevent accidental displacement of critical electrical parts.

The pseudo-balanced construction of Fig. 2 permits the use of a balanced transmission-line tank circuit, eliminating the necessity for a coaxial structure for this doubler stage. The radio-frequency power output of the first doubler is delivered next to a type *6161* tube operating in a doubler circuit of the grounded-grid type. The output power of this stage is approximately 40 watts and is utilized to drive a similar final doubler, again employing a type *6161* tube, and yielding a power output of 60 watts. These latter stages employ transmission-line types of circuits, in which the tube is completely enclosed by the outer cavity structure. Proper tuning for these stages is indicated by use of a reflectometer in the input line of the final doubler, and also through the use of suitable grid and plate ammeters in the second and final doubler stages.

It is interesting to note here that a grounded-grid doubler converts a portion of the radio-frequency drive power to second-harmonic power in the output circuit of the stage. In this respect it is identical in performance with a grounded-grid amplifier. Whereas it is difficult to obtain a stable amplifier at these frequencies using a *6161* tube in a grounded-grid circuit without the use of neutralization, the doubler circuit is completely stable and reliable in performance, with no neutralization. These frequency multipliers operate at an anode potential of approximately 1500 volts, and utilize a suitable combination of cathode and grid bias so that variations in tubes will not result in excessive variations of the radio-frequency power output. In addition, removal of the radio-frequency drive power cannot cause operation of the tube beyond its normal ratings. All tuning controls for the multiplier stages are brought out to the front panel, where meters are provided to indicate normal operation and the peak of various tuning controls. The radio-frequency power delivered by the final doubler is fed through a reflectometer to the input circuit of the aural power amplifier.

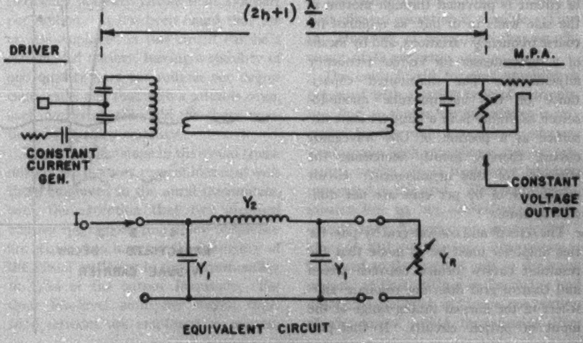
The aural power amplifier utilizes a type *6183* tetrode in a grid-screen separation type of circuit. The input circuit for this amplifier is of the conventiona coaxial type, having an electrical length of 3/4 wavelength. The output circuit for the tetrode consists of a waveguide beyond cutoff, with suitable mode suppressors and tuning stubs to furnish the proper load impedances for the amplifier tube. The output coupling circuit from this waveguide is a simple voltage probe, used as an impedance-matched transformer. For this waveguide type of circuit, Fig. 3, it can be shown that the resonant frequency can be closely calculated by assuming the circuit to consist of a transmission line operating in the transverse-electromagnetic mode, and having a width equal to 1/4 wavelength, a height equal to

Fig. 4. Constant-current to constant-voltage transformer, independent of changing load

$$\frac{V}{I_0}=\frac{Y_2/Y_R}{\frac{Y_1^2+2Y_1Y_2}{Y_R}-(Y_1+Y_2)}$$

at resonance $Y_1+Y_2=0$

$$\frac{V_1}{I_0}=\frac{Y_2}{Y_1^2+2Y_1Y_2}\text{ which is independent of }Y_R$$



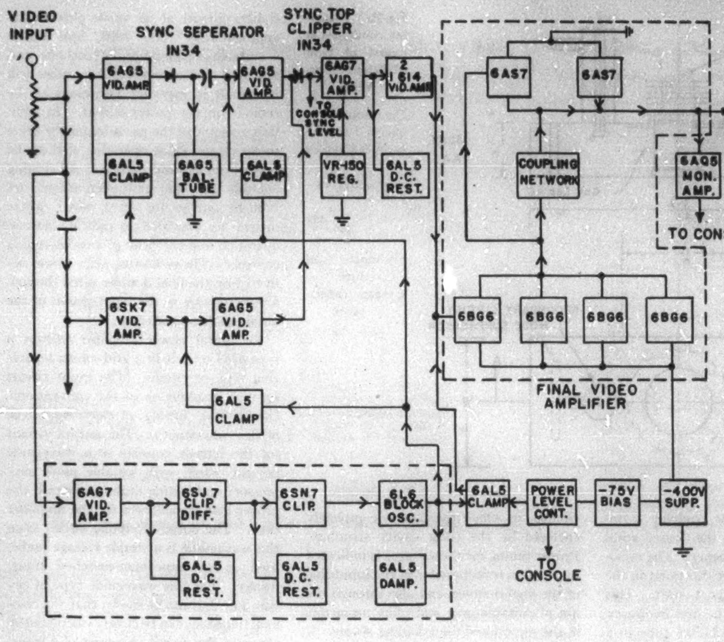

VIDEO INPUT
SYNC SEPERATOR IN34
SYNC TOP CLIPPER IN34
6AG5 VID. AMP.
6AG5 VID. AMP.
6AG7 VID. AMP.
2 1614 VID. AMP.
TO CONSOLE SYNC LEVEL
6AL5 CLAMP
6AG5 BAL. TUBE
6AL5 CLAMP
VR-150 REG.
6AL5 D.C. REST.
6AS7
6AS7
TO MPA GRID
COUPLING NETWORK
6AQ5 MON. AMP.
TO CONSOLE
6SK7 VID. AMP.
6AG5 VID. AMP.
6BG6
6BG6
6BG6
6BG6
6AL5 CLAMP
FINAL VIDEO AMPLIFIER
6AG7 VID. AMP.
6SJ7 CLIP. DIFF.
6SN7 CLIP.
6L6 BLOCK OSC.
6AL5 CLAMP
POWER LEVEL CONT.
-75V BIAS
-400V SUPP.
TO CONSOLE
6AL5 D.C. REST.
6AL5 D.C. REST.
6AL5 DAMP.

Fig. 5. Visual modulator

h, and a length equal to the width of the guide. This parallel-plane transmission line resonates with the tube capacitance at a frequency given by the implicit relation

$$\frac{1}{2\pi fC}=754\frac{fh}{c}\tan\frac{wf\pi}{c}$$

The function of the waveguide on either side of this strip of the resonator is to act as a reactive attenuator, suppressing radiation from the ends of the guide. Tuning for this waveguide type of circuit is provided through motion of the side walls in or out, as required for coarse frequency variations, and by means of a discondenser for vernier frequency adjustments. The distributed capacitance of the vacuum-tube anode-to-screen assembly is in a large measure absorbed as a portion of this waveguide circuit, thereby greatly increasing the efficiency of this arrangement; circuit efficiencies of 90 per cent are not difficult to obtain.

The screen- and control-grid by-pass for this amplifier must be so made that the resultant cavity formed between screen and control grid does not resonate anywhere in the normal tuning range of the input or output circuits. If this pre-

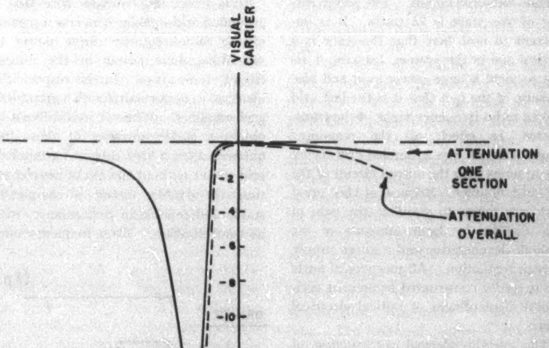
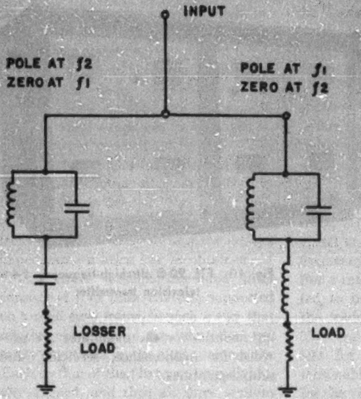


Fig. 6. (below). Attenuation characteristic vestigial-side-band filter



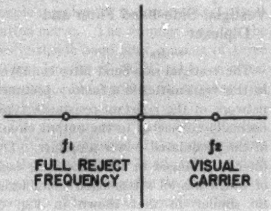


Fig. 7 (left). Low-frequency equivalent circuit of one section of vestigial-side-band filter

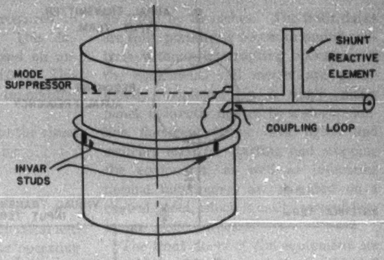


Fig. 8 (above). Typical cavity structure used in vestigial-side-band filter

caution is not observed, it is quite possible that the tube will oscillate, at a frequency determined by this screen-to-control-grid cavity. Through the suitable use of beryllium spring fingers it is possible to bypass the screen and control grids such that the cavity formed for this tube is resonant at 1,000 megacycles. This is well outside of the tuning band for either the input or output circuits of the amplifier. When the tube is utilized in this fashion, the low anode-to-cathode capacitance results in very stable operation.

This amplifier stage is arranged to operate with an anode potential of 3,800 volts, a screen potential of 500 volts, and a fixed bias of -60 volts. A combination of fixed bias and grid-leak bias is used, so that the tube cannot be damaged in the event of radio-frequency drive failure. When the aural transmitter is adjusted for maximum power output, the bandwidth is found to be in excess of 2 megacycles; hence, there is no possibility of producing frequency-modulation distortion or amplitude modulation of the aural signal. The power output of this stage is in excess of 600 watts, and the normal operating efficiency at 900 megacycles is 55 per cent. The cathode input circuit is arranged so that, by means of the directional coupler, a proper match on the drive line can be secured. The aural power-amplifier output is coupled to the main $1^{5}/_{8}$-inch transmission line through a directional coupler to permit monitoring of the output-power level and impedance matching. This power is then fed to the aural and visual diplexer to permit operation of the aural and visual transmitter on a common antenna without undue interaction.

General Description of Visual Transmitter

The visual-transmitter signal is, as mentioned above, originated in a crystal oscillator operating at 1/96 of the final output frequency. This crystal oscillator has been specially developed to secure the best possible stability with temperature, aging effects, and line-voltage variations. It is isolated from its load circuit by use of a buffer amplifier operating as a class *A* stage. Measurements made on this circuit show a long-time frequency stability better than 1/2 part per million. It has been found that the crystal employed in this circuit can be a commercial variety, having a stability of one quarter part per million per degree centigrade, and that with a suitable oven, such a crystal will yield the above-mentioned frequency stability.

The multiplier stage in the visual transmitters are in every respect identical with those employed in the aural transmitter, with the exception that two identical doubler stages operating at low frequency are required to increase the frequency of the visual oscillator to one corresponding to 1/24 of the output frequency. For these low-level multiplier stages, type-*1614* tetrodes are employed in conventional grid-driven circuits. The output of the final doubler, utilizing a type-*6161* tube, is fed to the visual-transmitter driver, which consists of a type-*6183* tetrode, again employed in a grid-driven separation type of circuit. This circuit is again in every way identical with that employed in the aural transmitter with the exception that the power level of operation is reduced, so that the output power of the driver is *350W*.

The use of a high-powered tube in this amplifier circuit enables one to obtain the following advantages of operation:

1. Large quantities of power available to obtain good regulation of drive for the visual-modulated power amplifier;
2. Good tube life for the driver circuits;
3. Uncriticalness of adjustment for the drive circuits.

The output power of this driver is delivered to the cathode-input circuit of the modulated power amplifier, which again utilizes a type-*6183* tetrode in a grid-screen separation circuit. The output cavity for this amplifier is of the double-tuned variety. This double tuning provides radio-frequency bandwidth of 5 megacycles, flat to within ±1/2 decibel. At the same time, a high load impedance can be presented to the plate-screen gap of the radio-frequency amplifier tube, so that efficient operation is possible. Thus this amplifier, at black level, has an efficiency of 43 per cent.

The control grid for this amplifier is bypassed with a radial type of transmission line, so that a low impedance is presented for the radio-frequencies, whereas a relatively high impedance can be presented to the visual modulator which is connected to this grid. This radial transmission line has an equivalent high-frequency radio-frequency impedance of 1/2 ohm, at 650 megacycles. The screen is bypassed to the control grid in the same manner as that utilized in the driver and aural final amplifier, and forms one portion of the above radial transmission line.

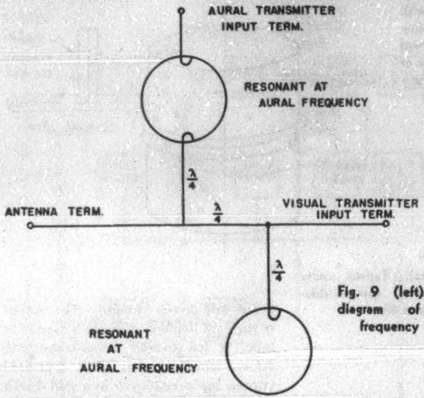


Fig. 9 (left). Schematic diagram of ultrahigh-frequency diplexer

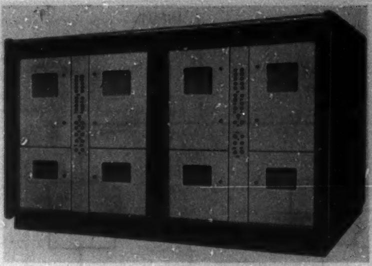


Fig. 10. FTL 20-B ultrahigh-frequency 1-kw television transmitter

The cathode circuit of the modulated power amplifier is damped by means of a radio-frequency load, which dissipates the power not utilized in the driving of the final visual amplifier. The dummy load for this purpose consists of a length of *RG14U* cable terminated in *RG21U* cable. This load is directly connected to the cathode circuit, where its position is adjustable to allow the desired ratio of power split between the cathode and dummy-load circuit.

It is quite easily shown that this kind of operation is not necessary, and that, in its place, the desired regulation of the radio-frequency drive can be secured through use of a coupling section between driver and modulated power amplifier, consisting of an odd number of quarter wavelengths of transmission line, see Fig. 4. However, the sensitivity of this kind of structure to misadjustment is such that it was felt desirable to avoid these difficulties, and to secure operation which was relatively uncritical to the setting of the driver tuning controls. With the arrangement employed in this transmitter, modulation to within 5 per cent of complete carrier cutoff is easily obtained, with excellent linearity. In addition, because of the radio-frequency loading on the modulated amplifier input circuits, the stability is greatly improved. The peak radio-frequency power capacity of the modulated power amplifier is such that 1,200 watts of synchronous peak power is easily obtained. Proper tuning of the input and output circuits is facilitated through the use of directional couplers and suitable diode-detector indicating equipment.

The modulator signal for the grid of the final amplifier is obtained from the circuit shown in Fig. 5. This modulator is required to deliver a signal having the following characteristics:

1. 100 volts peak-to-peak, with −70 volts equivalent to black level.
2. The modulator load is 200 micromicrofarads, with a peak load current, at —40 volts, equal to 130 milliamperes.
3. At —110 volts, the load current is zero milliamperes.

The modulator employed is conventional in every respect and has been described in other papers, with the exception of the output stage. The output stage of the modulator utilizes a so-called shunt-regulated amplifier circuit. This circuit has the advantage of furnishing an extremely low equivalent-output impedance. Referring to Fig. 5, we see that four type-*6BG6* tubes are arranged in parallel to drive two type-*6AS7* tubes, in this shunt amplifier circuit. This combination of tubes has an equivalent internal impedance of 50 ohms, and is likewise easily capable of supplying the peak current of 130 milliamperes at a potential of −40 volts. Furthermore, inasmuch as the load resistor required for this circuit is small, the equivalent efficiency is excellent. Thus for 5 megacycles video bandwidth, our power consumption at white picture is 300 milliamperes, at a potential of 400 volts. The modulator is insensitive to load impedance to such an extent that the output load capacitance can be increased to 400 micromicrofarads with no noticeable deterioration of picture quality. The modulator output circuit contains a considerable quantity of negative feedback as a result of the action of *R1*, see Fig. 5, so that the linearity of this arrangement is excellent. The total distortion of the modulator from 5 percent to 75-per-cent, peak synchronous bias is less than 5 per cent. Monitoring facilities for the modulator are provided for maintenance, servicing, and setup operations.

Vestigial-Side-band Filter and Diplexer

The vestigial-side-band filter employed in this transmitter is a factory pretuned network of the constant-resistance type, normally connected to the output circuit of the modulated power amplifier. The filter is arranged in three sections, each of which has an attenuation characteristic similar to that shown in Fig. 6. These three sections are cascaded, and the frequency of maximum attenuation is so set relative to the frequency of minimum attenuation that the over-all characteristic is similar to that shown in Fig. 6. To obtain this characteristic, a network having the low-frequency equivalent shown in Fig. 7 is required.

A calculation of the network parameters reveals that, for operation at 890 megacycles, if the insertion loss of the filter at visual carrier is to be less than 0.5 decibel the unloaded Q's of the reactive elements must be in the vicinity of 20,000. To realize circuit elements having this quality of performance it was necessary to plan on the use of resonant cavities. The most promising type of cavity appeared to be of cylindrical shape, operating in the TE_{111} mode. The theoretical Q for such a cavity, constructed of copper, and of height nearly equal to diameter, was 50,000 at a frequency of 650 megacycles. This form of structure was adapted for use in the filter and diplexer. The cavities, as finally constructed, are made of copper-clad invar sheets, and are deep drawn to size. Tuning is accomplished by splitting the structure in half, perpendicular to the cavity axis, see Fig. 8, and moving the halves with respect to each other. This gives a tuning variation of 80 megacycles per inch. The cavity halves are held together with threaded invar rods, to insure temperature stability.

The proper impedance is realized on coupling the cavity to the drive line through use of coupling loops of suitable size. A mode suppressor is employed to prevent the generation of the TE_{111} mode at right angles to the desired mode. The shunt-reactive element shown in Fig. 7 is constructed of $1^4/_8$-inch air line, and is arranged at a suitable distance from the main cavity so that the desired impedance variation can be obtained.

The entire vestigial-side-band filter consists of 6 of these cavities, supported on a rigid steel frame, in such a way that relative expansion of the frame does not apply any strain on the filter components. Each section of the filter can be individually aligned, and then all three sections couple together for the desired final attenuation curve. The dummy load for the vestigial-side-band filter consists of three 50-foot lengths of *RG8U* cable, on a reel structure. No cooling is required for this load.

The diplexer, see Fig. 9, for equivalent circuit is constructed of the same cavities as the vestigial-side-band filter, and is mounted and arranged in the same general manner. The aural insertion loss for such a diplexer is given by

$$\frac{P_{\text{in}}}{P_{\text{out}}}=\left(1+\frac{Q_L}{Q_0}\right)^2$$

where Q_0 is the unloaded cavity Q, and Q_L is the loaded Q of the cavity. For this diplexer, it was felt desirable to have the insertion loss less than 1 decibel whereas the loaded bandwidth was required to be at least 0.6 megacycle. This dictated a loaded Q of 1,000, and an unloaded Q of approximately 30,000. This was well within the theoretical capabilities of the cavity structure.

The crosstalk for the aural to the visual transmitter is similarly given by

$$\frac{P_{\text{visual}}}{P_{\text{load}}}\cong\left(\frac{1}{2EQ}\right)^2$$

where E is the interval between aural and visual carriers, divided by the operating frequency, and Q is as in the foregoing. For a ratio of −24 decibels, Q is calculated to be approximately 1,000, that is, the loaded Q is calculated to be 1,000. This value gives satisfactory insertion loss for the aural transmitter and a reasonable value of crosstalk. The loss for the visual line in the diplexer is generally negligible.

The impedance match on the visual line was easily made to have a voltage standing-wave ratio less than 1.1 and the aural line less than 1.5. The transmission line used to connect the cavities is $1^5/_8$-inch diameter. Even with 10 kw of combined aural and visual power, the heating in the short length of line employed is negligible.

Cabinet, Cooling, and Control-Circuit Arrangement

The entire transmitter is enclosed in a cabinet for shielding, structural support, and convenience in shipping, see Fig. 10. The aural and visual transmitters each occupy one bay approximately 7 feet by 5 feet by 31 inches. The front doors are split, permitting access to tuning controls without interrupting operation of the equipment. All meters are prominently displayed on a single meter panel, indirectly lighted by a single slimline florescent lamp. The lights and switches used for starting and stopping the equipment, as well as indicating normal functioning, are mounted on a central panel, which is easily removed for maintenance purposes.

The front doors of the equipment are provided with large windows, shielded with perforated metal, so that the interior of the transmitter is clearly visible when operating. The rear doors of the transmitter are of the sliding-panel type, two to a single bay. Thus, no space is required in the rear of the transmitter for swinging doors. These sliding doors can be completely removed from their raceways, thereby giving excellent access to the rear of the transmitter. Cooling for each bay is provided by a single blower-filter combination yielding 600 cubic feet per minute at a 3-inch head. A centralized duct system is used to distribute this air where needed. The control circuits are of such a nature that automatic starting and stopping of the equipment is possible through the use of a single switch. Power supply requirements for standard 1-kw plate black picture output are approximately 10 kva, 208 volts, 3-phase, 4-wire, 50 to 60 cycles at 90-per-cent power factor.

◆

No Discussion

Block-Diagram Solutions for Vacuum-Tube Circuits

T. M. STOUT
ASSOCIATE MEMBER AIEE

MANY AUTHORS[1–7] of textbooks on electronics begin their discussion of feedback amplifiers with the block diagram given by Black.[8] The general gain equation

$$A'=\frac{A}{1-A\beta} \tag{1}$$

in which A is the gain without feedback and β is the feedback fraction, is then derived, and the beneficial effects of feedback are discussed. In dealing with specific circuits, however, it is a common practice to write and solve the circuit equations, and then to manipulate the resulting gain equation into a form in which A and β can be identified. In some cases, expressions for A and β are simply stated and verified by substitution into the general equation.

In this paper, some simple amplifiers and oscillators are studied using block-diagram methods which have been applied to feedback-control systems[9] and electric circuit problems.[10] Using these methods, a block diagram based on the equivalent circuit is first drawn. This diagram is then reduced by a series of block transformations until the A and β portions of the system have been isolated.* By this procedure, the amplifier gain can be obtained in a logical and satis-

---

Paper 53-269, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted August 5, 1952; made available for printing April 20, 1953.

T. M. Stout is with the University of Washington; Seattle, Wash.

* These diagrams resemble the signal flow graphs, which were described by S. J. Mason at the Institute of Radio Engineers National Convention, New York, N. Y., March 20, 1951.

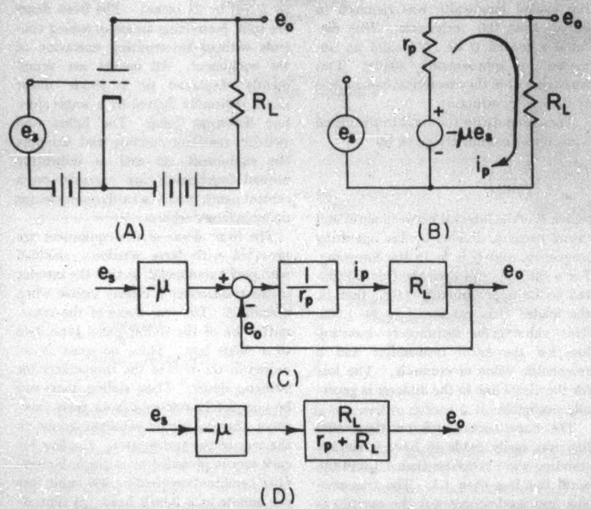


Fig. 1. Triode-amplifier stage with resistive load, equivalent circuit, and block-diagram solution

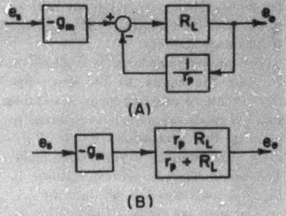


Fig. 2. Alternative method for reducing triode-amplifier block diagram

fying way without formal solution of the system equations.

With the various parts of the system clearly defined, the Routh and Nyquist stability criteria can be used with confidence. Knowing that a complete block diagram can be derived for vacuum tube circuits, the electronics specialist may be encouraged to use the highly-developed and very convenient graphical techniques of the servomechanism designer: standardized curves and templates for the common gain factors, attenuation-phase plots and the associated Nichols' charts for determining allowable gains, and a variety of compensation methods and design charts.[11–13]

The general procedures and a detailed description of the rules for block-diagram transformation are given in previous papers.[9,10] With the exception of the general feedback equation given, these rules are based on the associative and distributive laws of algebra, and can be understood from the examples given in this paper.

Amplifiers Without Feedback

In addition to providing an introduction to block-diagram methods, study of some simple amplifiers without feedback will furnish a basis for later comparison of these amplifiers with feedback amplifiers. Consider first the single-stage triode amplifier with resistive load, shown in Fig. 1(A). In drawing the equivalent circuit, Fig. 1(B), the tube is represented as a voltage generator in series with the plate resistance.

The block diagram, given in Fig. 1(C), is based on the circuit equations

$$i_p = \frac{1}{r_p}(-\mu e_s - e_o) \qquad (2)$$

$$e_o = i_p R_L \qquad (3)$$

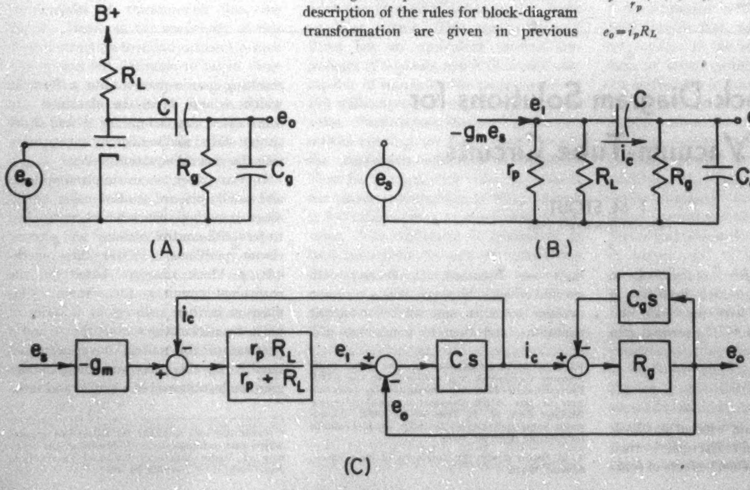


Fig. 3. Resistance-capacitance coupled amplifier stage with equivalent circuit and block diagram

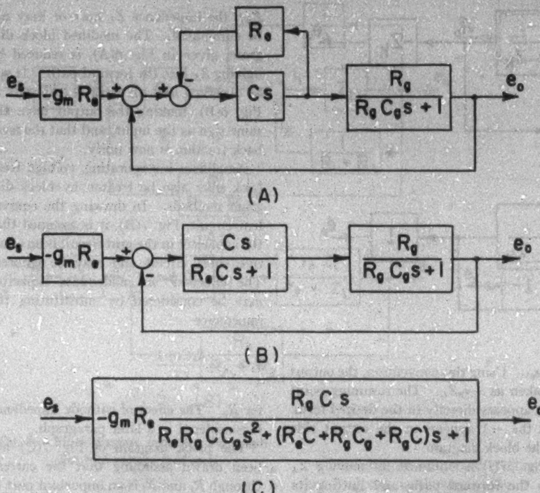


Fig. 4. Block-diagram reduction for resistance-capacitance coupled amplifier stage

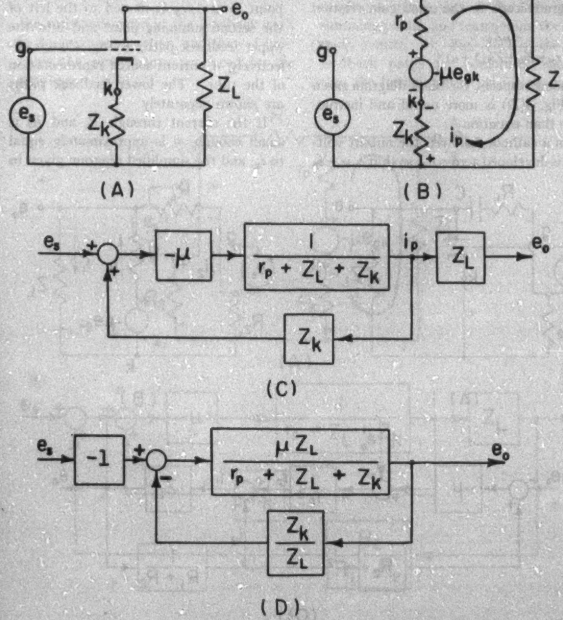


Fig. 5. Amplifier with current feedback and block-diagram solution

In reducing the block diagram, the general formula for feedback systems is applied, taking the forward gain as R_L/r_p and the feedback fraction as -1. (The negative character of the feedback is indicated by the minus sign at the summing point. Following the practice used in servomechanism analysis, the negative sign could be absorbed in the general feedback formula, equation 1, by writing the denominator with a positive sign; β would then be considered as normally positive.) As expected, the gain of the amplifier is

$$\frac{e_o}{e_s} = -\mu \frac{R_L}{r_p + R_L} \tag{4}$$

represented in block form in Fig. 1(D).

This block diagram can be reduced in another way. Since none of the circuit equations is disturbed, the small block containing $1/r_p$ can be moved to the left of the summing point and into the feedback path. Then, replacing μ/r_p by g_m, the diagram of Fig. 2(A) is obtained. From this diagram, another equivalent circuit can be drawn in which the tube is represented as a current generator supplying a parallel combination of r_p and R_L. Thus the equivalent circuit which is usually derived by means of Norton's theorem can be established by block-diagram transformation. Application of the general formula for feedback systems to Fig. 2(A) produces the final result of Fig. 2(B), showing that the gain is g_m multiplied by the equivalent resistance of r_p and R_L, as anticipated.

The voltage gain of a resistance-capacitance coupled amplifier can be obtained by a combination of procedures developed for ladder networks[10] with the ideas used in the preceding analysis. The circuit considered, and an equivalent circuit based on a current-source representation of the tube are given in Figs. 3(A) and 3(B). The block diagram, Fig. 3(C), correctly indicates the necessary equations.

Fig. 4(A), in which R_e represents the equivalent resistance of r_p and R_L in parallel, is the result of a partial reduction of the original block diagram. Three changes are incorporated:

1. The feedback loop, including R_g and C_g, has been reduced to a single block;
2. The block containing R_e has been moved to the left of the first summing point and into the upper feedback path;
3. The order of the first two summing points has been changed.

In the next diagram, Fig. 4(B), the inner feedback loop containing R_e and C has been reduced to a single block. The final block diagram, Fig. 4(C), is obtained by first combining the two blocks

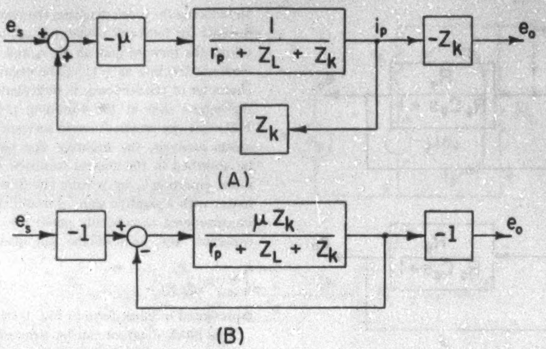


Fig. 6. Cathode-follower block diagram

in the forward path and then applying the general rule for feedback systems.

Recalling that s may be replaced by $j\omega$ to obtain the gain and phase shift as functions of frequency, it will be observed that the final block diagram, Fig. 4(C), contains all the factors necessary to account for the decrease in gain at low and high frequencies. In fact, it also includes the information required to determine the step or pulse response by operational methods.

Feedback Amplifiers

Having examined two representative amplifier circuits which appear to include numerous feedback loops, it may seem superfluous to devote space specifically to feedback amplifiers. However, the block diagram method demonstrates nicely some of the main properties of feedback amplifiers. It shows clearly, for example, that the tube parameters (μ, g_m, and r_p) occur entirely within the principal feedback loop and, to be more specific, in the forward part of the loop. It identifies plainly which passive elements enter into the transfer functions of the forward and feedback paths.

The first example considered is the case of current feedback from an element in the cathode circuit, shown in Fig. 5. With the sign conventions employed, the feedback appears to be positive, as indicated in Fig. 5(C). However, moving the negative sign associated with the amplification factor to the left of the summing point brings about, in addition to the -1, indicating the sign reversal of the amplifier as a whole, a change in sign at the summing point. The equivalent circuit could also be drawn with current leaving the cathode and a reversed voltage equal to $+\mu e_{gk}$. Using this convention, the output is taken as $-i_pZ_L$. The summing point now appears directly in the desired form, and the -1 appears at the output side of the block diagram.

Fig. 5(D) is obtained by moving Z_L into the forward path, and putting its reciprocal into the feedback path. The large block in the forward path is now the gain-without-feedback, and the block in the feedback path is the feedback fraction. Further reduction of the block diagram leads to the usual gain relation

$$\frac{e_o}{e_s}=\frac{-\mu Z_L}{r_p+Z_L+Z_K(1+\mu)} \tag{5}$$

In some respects, the block diagram given in Fig. 5(D) is more useful and instructive than equation 5.

In a cathode follower, the output voltage is developed across Z_K, so that $e_o=-i_pZ_K$; the impedance Z_L may or may not be eliminated. The modified block diagram, given in Fig. 6(A), is reduced by moving Z_K into the forward path. It will be observed that an extra -1 appears in Fig. 6(B), making the output have the same sign as the input, and that the feedback fraction is now unity.

Amplifiers incorporating voltage feedback may also be treated by block-diagram methods. In drawing the equivalent circuit, Fig. 7(B), it is assumed that the capacitor in the grid circuit is an effective short circuit and may be ignored. The effect of an inadequate capacitor may be considered by substituting the impedance

$$R_1+\frac{1}{Cs}=\frac{R_1Cs+1}{Cs}$$

for R_1. The effect of cathode impedance is considered in a later paragraph.

The block diagram of Fig. 7(C) has been drawn assuming that the current through R_1 and R_2 is an important part of the total tube current. To obtain the next diagram, Fig. 8(A), several steps are taken. As before, the negative sign associated with the amplification factor is moved to the left of the first summing point, and $1/r_p$ is moved to the left of the second summing point and into the upper feedback path, giving what is effectively a current-source representation of the tube. The lower feedback paths are shown separately.

If the current through R_1 and R_2 is small enough, i_1 is approximately equal to i_p, and the simplified diagram given in

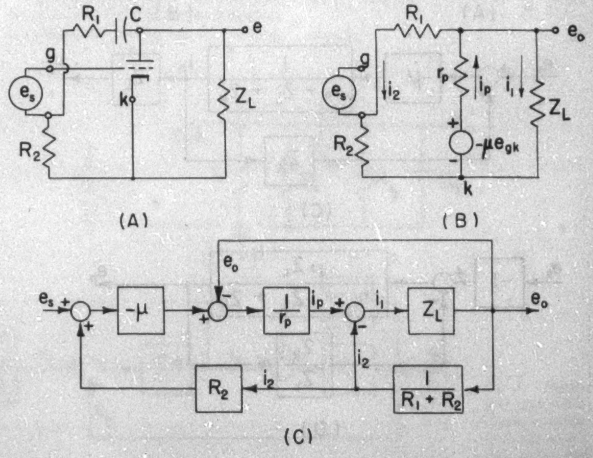


Fig. 7. Amplifier with voltage feedback and block diagram

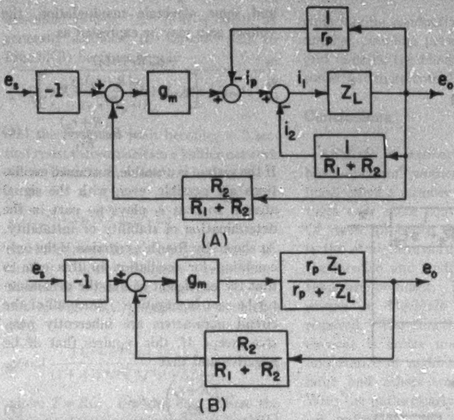


Fig. 8. Block-diagram reduction for amplifier with voltage feedback

Fig. 8(B) can be drawn. The blocks in the forward path represent the gain-without-feedback

$$\frac{g_m r_p Z_L}{r_p+Z_L}=\frac{\mu Z_L}{r_p+Z_L}$$

the resistance ratio in the feedback path is the feedback fraction, and the -1 indicates the sign change produced by the amplifier.

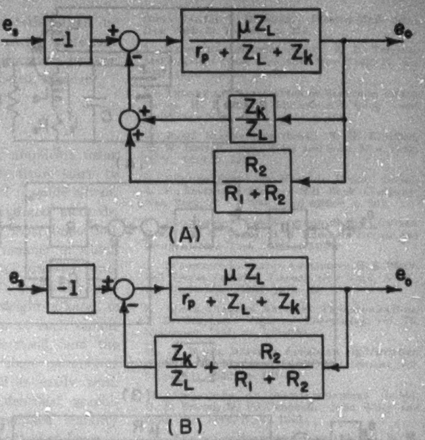


Fig. 10. Block-diagram reduction for amplifier with combined current - voltage feedback

Compound or combined voltage-current feedback is shown in Fig. 9. For simplicity, the capacitor in the grid circuit is not included, and it is assumed at the start that the current through R_1 and R_2 is negligible. Fig. 10(A) is obtained by moving Z_L into the inner loop, moving a minus sign to the left of the summing points, and rearranging the summing points. In Fig. 10(B), the two feedback paths are combined, showing that the feedback fraction for compound feedback is simply the sum of the feedback fractions for the two types of feedback separately.

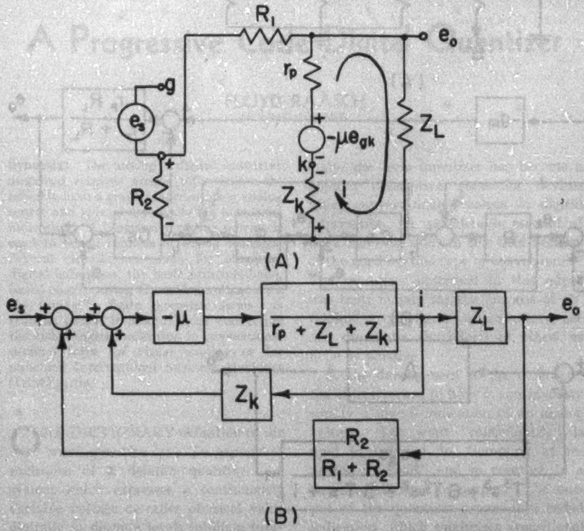


Fig. 9. Amplifier with combined current-voltage feedback, and block diagram

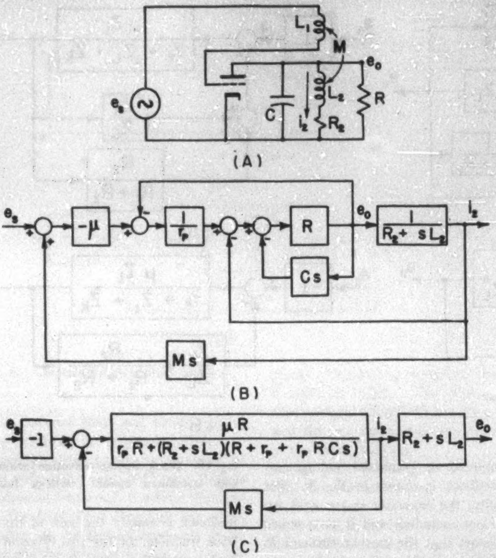
Grid-plate capacitance and other parasitic parameters can be considered by appropriate modifications of the block diagrams. Multistage amplifiers, with and without feedback, are treated by obvious extensions of the ideas presented here.

Oscillators

Oscillators are often treated as unstable feedback systems, a procedure which is neatly justified by block-diagram methods. Consider, for example, the tuned-plate oscillator shown in Fig. 11(A). Here the similarity of this oscillator circuit to a feedback amplifier is emphasized by showing a separate signal voltage in the grid circuit. Grid current is neglected and it is assumed that any load may be represented as a shunt resistance, R; the mutual inductance may be either positive or negative, but it is assumed initially that the voltage induced in the grid circuit adds to the signal voltage.

The block diagram, Fig. 11(B), represents correctly all of the required circuit equations and follows the pattern developed in studying feedback amplifiers. After several routine steps, the diagram of Fig. 11(C) is obtained. This diagram includes the -1 which is indicative of the sign change associated with a single stage of amplification, and the signs at the summing point indicate negative feedback.

After one more block transformation



L_1
M
e_0
e_s
C
i_2
L_2
R_2
R
(A)
e_s
$-\mu$
$\frac{1}{r_p}$
R
e_0
$\frac{1}{R_2+sL_2}$
i_2
Cs
Ms
(B)
e_s
-1
$\frac{\mu R}{r_pR+(R_2+sL_2)(R+r_p+r_pRCs)}$
i_2
R_2+sL_2
e_0
Ms
(C)

Fig. 11 (above). Tuned-plate oscillator and block diagrams

and some algebraic manipulation, the voltage gain may be expressed as

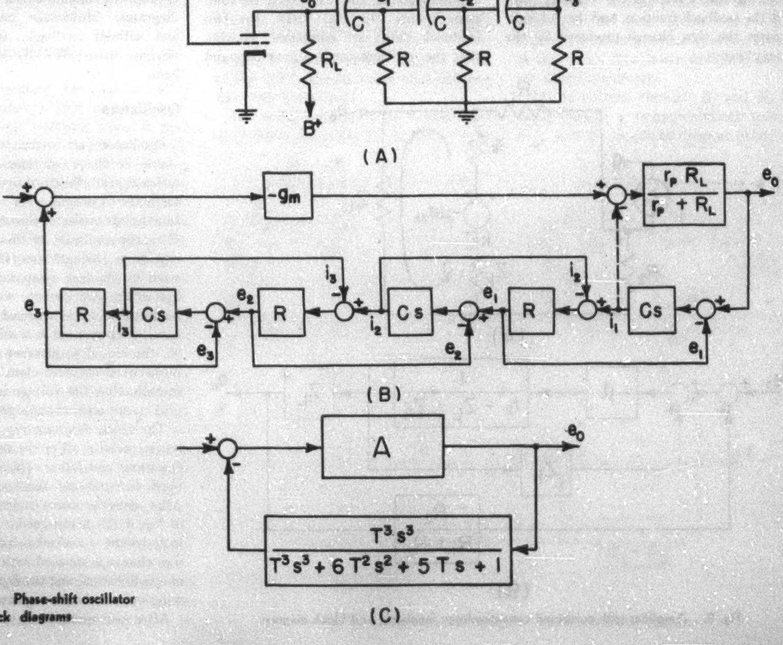
$$\frac{e_0}{e_s}=\frac{-g_m(R_2+sL_2)}{s^2L_2C+s\left[g_mM+L_2\frac{(R+r_p)}{Rr_p}+R_2C\right]+1+R_2\frac{(R+r_p)}{Rr_p}} \qquad (6)$$

If the system is unstable, sustained oscillations are possible, even with the signal absent, so that e_s plays no part in the determination of stability or instability. As shown by Routh's criterion,[13] the only condition for oscillation in this case is that the coefficient of s in the denominator be zero or negative. Since all of the circuit parameters are inherently positive except M, this requires that M be negative and that

$$|M|\geq\frac{L_2\left(\frac{R+r_p}{Rr_p}\right)+R_2C}{g_m} \qquad (7)$$

Equation 7 may also be used to estimate the transconductance needed to produce oscillations.

A similar treatment of the resistance-capacitance phase-shift oscillator[14] is



i_1
i_2
i_3
e_0
e_1
e_2
e_3
C
C
C
R_L
R
R
R
B^+
(A)
$-g_m$
$\frac{r_p R_L}{r_p + R_L}$
e_0
e_3
R
Cs
i_3
e_2
R
Cs
i_2
e_1
R
i_1
Cs
(B)
A
e_0
$\frac{T^3s^3}{T^3s^3+6T^2s^2+5Ts+1}$
(C)

Fig. 12 (right). Phase-shift oscillator block diagrams

presented in Fig. 12. If i_1 is small, as generally assumed, the forward path in Fig. 12(B) becomes simply

$$A = g_m \frac{r_p R_L}{r_p + R_L} = \mu \frac{R_L}{r_p + R_L} \tag{8}$$

and the feedback path becomes a 3-section resistance-capacitance ladder network whose transfer function

$$\frac{e_{\text{out}}}{e_{\text{in}}} = \frac{T^3 s^3}{T^3 s^3 + 6T^2 s^2 + 5Ts + 1} \tag{9}$$

can be derived in a variety of ways, including block diagram reduction.[10] Viewed as a feedback amplifier, the system of Fig. 12(C) would have a voltage gain

$$A' = A \frac{T^3 s^3 + 6T^2 s^2 + 5Ts + 1}{(1+A)T^3 s^3 + 6T^2 s^2 + 5Ts + 1} \tag{10}$$

where $T = RC$. For this case, where the denominator is a cubic in s with positive coefficients, Routh's criterion gives the necessary condition for oscillation as

$$T^3(1+A) \geqq (6T^2)(5T) \tag{11}$$

or

$$A \geqq 29$$

as is well known.[14]

Other oscillator circuits, such as the Hartley, Colpitts, or Wien-bridge, can be analyzed by similar methods. For circuits having conductive, rather than inductive, coupling between the plate and grid circuits, the block diagrams become somewhat more complicated, however.

Conclusions

Analysis of feedback amplifiers using block-diagram methods from start to finish offers a number of possible advantages over more conventional methods. Of these, probably the most important is the straightforward manner in which the forward and feedback transfer functions are obtained. No particular prescience or algebraic sleight-of-hand is required. The function of each circuit element is better understood, and the importance of making some parameters large and others small is easily seen. With the paths clearly identified, graphical techniques and standard stability criteria can be applied with confidence.

Block diagram methods emphasize the similarity between feedback amplifiers and oscillators, encouraging joint use of techniques for analysis and synthesis.

References

1. ELECTRON-TUBE CIRCUITS (book), **S. Seeley.** McGraw-Hill Book Company, New York, N. Y., 1950, p. 82.

2. ELECTRONIC FUNDAMENTALS AND APPLICATIONS (book), **J. D. Ryder.** Prentice-Hall, Inc., New York, N. Y., 1950, p. 272.

3. RADIO ENGINEERING (book), **F. E. Terman.** McGraw-Hill Book Company, New York, N. Y., 1947, third edition, p. 311.

4. THEORY AND APPLICATION OF ELECTRON TUBES (book), **H. J. Reich.** McGraw-Hill Book Company, New York, N. Y., 1944, p. 198.

5. BASIC ELECTRONICS (book), **R. G. Kloeffler, M. W. Horrell.** John Wiley and Sons, New York, N. Y., 1949, p. 191.

6. FUNDAMENTALS OF VACUUM TUBES (book), **A. V. Eastman.** McGraw-Hill Book Company, New York, N. Y., 1949, third edition, p. 361.

7. FUNDAMENTALS OF ELECTRONICS AND CONTROL (book), **M. G. Young, H. S. Bueche.** Harper and Brothers, New York, 1952, p. 340.

8. STABILIZED FEEDBACK AMPLIFIERS, **H. S. Black.** *Bell System Technical Journal*, New York, N. Y., vol. 13, January, 1934, pp. 1–18.

9. BLOCK DIAGRAM NETWORK TRANSFORMATION, **T. D. Graybeal.** *Electrical Engineering*, vol. 70, November, 1951, pp. 985–990.

10. A BLOCK DIAGRAM APPROACH TO NETWORK ANALYSIS, **T. M. Stout.** *AIEE Transactions*, vol. 71, part II, 1952, pp. 255–60.

11. PRINCIPLES OF SERVOMECHANISMS (book), **G. S. Brown, D. P. Campbell.** John Wiley and Sons, New York, N. Y., 1948.

12. THEORY OF SERVOMECHANISMS (book), **H. M. James, N. B. Nichols, R. S. Phillips.** MIT Radiation Laboratory Series, vol. 25, McGraw-Hill Book Company, New York, N. Y., 1947.

13. SERVOMECHANISMS AND REGULATING SYSTEM DESIGN (book), **H. Chestnut, R. W. Mayer.** John Wiley and Sons, New York, N. Y., 1951.

14. PHASE-SHIFT OSCILLATORS, **E. L. Ginzton, L. M. Hollingsworth.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 29, February, 1941, pp. 43–49.

◆

No Discussion

A Progressive Code Digital Quantizer

FLOYD RAASCH
NONMEMBER AIEE

Synopsis: The analog-to-digital converter described receives analog information obtainable from a graph reader or other analog source and gives a corresponding numerical indication. A progressive coding scheme is used to lessen chances of wrong readings. Several circuits are given for obtaining digital indication, the basic principle being based on comparing the analog voltage with the sum of a finite geometric series. A readout circuit is described for converting the coded digital indication to conventional decimal form for visual reading or for punching International Business Machines (IBM) cards.

ONE DICTIONARY definition of the word quantize is: "To express as multiples of a definite quantity." A system which expresses a continuously variable voltage or other physical value digitally in discrete levels fulfills a function which fits this definition. Accordingly, the term quantizer has become a popular descriptive name for devices which convert analog values to digital information. A considerable amount of work has been done in this field because of the need for this type of conversion.[1–7] The quantizer described in this paper was built mainly for the purpose of obtaining a quantizer for laboratory use and duplicates the efforts of others as little as possible.

In the development of the quantizer, the objective was to have it continuously supply a digital indication of an analog voltage. The word "continuously" is used in contrast to "sampling at repeated intervals" and in contrast to a "sequential" form of output. The output of the quantizer proper is a coded indication, which minimizes transitional difficulties between those adjacent incremental values which, in more conventional number systems, involve the change of more than one digit.

A readout circuit converts the coded reading to conventional decimal readings to make the output easily intelligible for general purposes. This readout circuit operates both an indicator panel and an IBM type-*513* reproducing punch.

Need for an Appropriate Number System

Since this instrument was designed to supply an indicator continuously, it was necessary to design the equipment so that a change from one digital value to the next adjacent incremental value is not accompanied by a momentary ambiguous indication, or a wrong value. In

---

Paper 53-287, recommended by the AIEE Computing Devices and Instruments and Measurements Committees and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 20, 1953; made available for printing April 29, 1953.

FLOYD RAASCH is with the Minneapolis-Honeywell Regulator Company, Minneapolis, Minn.

Table I. Progressive Decimal Number System

Progressive Code	Decimal System	Progressive Code	Decimal System	Progressive Code	Decimal System
000	000	060	060	⋮	⋮
001	001	⋮	⋮	210	219
002	002	069	069	220	220
003	003	079	070	⋮	⋮
004	004	⋮	⋮	229	229
005	005	070	079	239	230
006	006	080	080	⋮	⋮
007	007	⋮	⋮	230	239
008	008	089	089	240	240
009	009	099	090	⋮	⋮
019	010	⋮	⋮	249	249
018	011	090	099	259	250
⋮	⋮	190	100	⋮	⋮
012	017	⋮	⋮	250	259
011	018	199	109	260	260
010	019	189	110	⋮	⋮
020	020	⋮	⋮	269	269
⋮	⋮	180	119	279	270
029	029	170	120	⋮	⋮
039	030	⋮	⋮	270	279
⋮	⋮	179	129	280	280
030	039	169	130	⋮	⋮
040	040	⋮	⋮	289	289
⋮	⋮	100	199	299	290
049	049	200	200	⋮	⋮
059	050	⋮	⋮	290	299
⋮	⋮	209	209	390	300
050	059	219	210		

the more commonly used number systems, changing a value representation by a difference of one unit is often accomplished by changing several or all significant digits in that number. In some types of devices for giving digital indications of continuously variable values, some time lag exists between changes in the separate digital positions. In other devices, there is a simultaneous instant of ambiguity in all digital positions undergoing change. Where automatic means to record a reading at a given instant is to be employed, conventional numbering systems present a handicap if a reading is taken at the instant of ambiguity or erroneous indication.

Number System Employed

To overcome such a handicap, the quantizer proper converts analog voltage values into a digital indication such that any two values differing by one increment are separated by a change of only one digit. The usual decimal numbering system is rearranged to represent progressively changing values differing by one unit in such a way that in all cases only one digit is changed to represent the next higher or lower value. This new arrangement is compared with the common decimal system in Table I. After the digit in a given position has progressed from zero to its maximum value, the numbering order of the digit values in that position is reversed for the next succession of representations. In a system of tens, after a digit has progressed from zero to nine, the nine is retained to represent the value ordinarily represented by zero in the conventional numbering system. The sequence of digits is then followed in reversed order until the extreme, now represented by zero, is reached. The zero is retained to indicate minimum value, and the digits again follow in the order of the normal decimal system.

This general numbering scheme of course is usable not only for a system of tens in particular, but for numbering systems of any radix in general. When we apply the scheme to the minimum possible radix, a system based on two, a specific case is reached which has been called the Gray code or the cyclic progression code. Table II compares the Gray code with a conventional binary arithmetic system, and Fig. 1 shows this graphically.

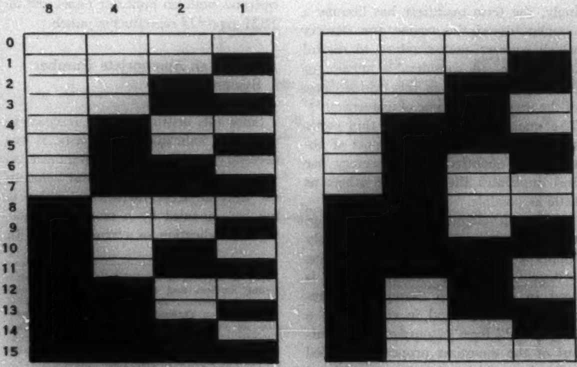


Fig. 1. Progressive binary number system. Left, binary code. Right, progressive code

Table II. Progressive Binary Number System

Progressive Code	Binary System	Decimal System
0000	0000	0
0001	0001	1
0011	0010	2
0010	0011	3
0110	0100	4
0111	0101	5
0101	0110	6
0100	0111	7
1100	1000	8
1101	1001	9
1111	1010	10
1110	1011	11
1010	1100	12
1011	1101	13
1001	1110	14
1000	1111	15
11000	10000	16
11001	10001	17
11011	10010	18
11010	10011	19
11110	10100	20
11111	10101	21
11101	10110	22
11100	10111	23

In all these progressive systems, the test of whether any digit represents actual value or whether it represents complement minus one is to note if the sum of the higher order digits is even or odd. If the sum of the higher order digits is even, the digit in question represents actual value. If the sum of the higher order digits is odd, the digit represents complement minus one. For example, if a progressive decimal number is expressed 126, the number can be converted to conventional decimal notation as follows:

1—Preceded by even sum (all zeros), therefore remains 1
2—Preceded by odd sum (one), therefore becomes $9-2=7$
6—Preceded by odd sum (three), therefore becomes $9-6=3$

The decimal number represented then is 173.

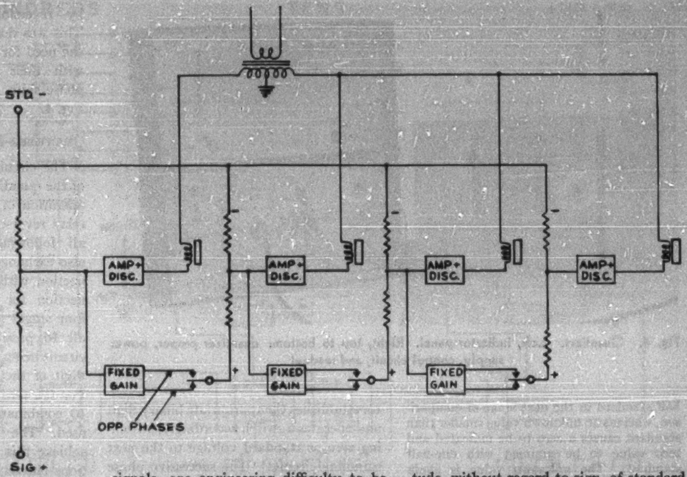
If a Gray code (progressive binary) number is expressed as 11011, its conversion to conventional binary is as follows:

1—Sum of higher order digits even (all zeros), therefore remains 1
1—Sum of higher order digits odd (one), therefore becomes 0
0—Sum of higher order digits even (two), therefore remains 0
1—Sum of higher order digits even (two), therefore remains 1
1—Sum of higher order digits odd (three), therefore becomes 0

The binary number represented than is 10010.

The quantizer described here combines the progressive binary systems and the progressive decimal systems. A group of four progressive binary digits represents a progressive decimal digit. No more than one decimal digit changes between any two incremental changes and within that coded decimal digit only one progressive binary digit changes between

Fig. 2. Possible progressive code digital quantizer



any two values. Thus only 1 out of 12 digits in the over-all system changes at a time.

For purposes of easy reading and interpretation, a translating circuit translates this number system to a system of conventional notation. If more than one digit in the conventional indication changes between two adjacent values, this readout circuit causes these digits to change simultaneously in the translating process.

Possible Quantizers

As an approach to this problem, a quantizer was visualized in the form of a network of nonlinear elements. Such a network is possible with the voltage polarities or phases at certain points forming a digital indication of the voltage applied to the network. For example, it is theoretically possible to connect a network of squaring and square rooting devices in such a manner that the network serves as a quantizer.

Of course, nonlinear responses are not confined to curvilinear functions. Nonlinear functions also include broken-line responses and discontinuous straight-line functions. A broken straight line or a discontinuous straight line suggests a circuit which includes the operation of a switch or relay. A possible quantizer employing amplifiers and relays is shown in Fig. 2.

This circuitry gives a digital output in the form of a progressive code. If a quantizer is made in this manner for a-c signals, one engineering difficulty to be experienced is that of obtaining a good balance after the signal has experienced the cumulative distortion and phase shift of several stages. In this circuit, a signal input is compared to a reference in the first stage. When the difference between these two values changes sign, the indicator changes but the magnitude of the difference is fed to the following stage without reversal. A fixed-gain amplifier with a reversing relay to maintain constant polarity of output with reversal of input polarity is the nonlinear device which, in this case, gives a broken line characteristic. In the second stage, the magnitude, without regard to sign, of standard magnitude-minus-signal magnitude for that stage is compared with half standard. When this difference changes polarity, the second indicator changes indication; the magnitude of the difference is again fed to the following stage without reversal.

In Fig. 3 is shown an analog-to-digital converter which gives an indication in ordinary binary form in the following manner:

An unknown signal value is compared to a standard value. If the unknown is greater than standard a binary digit one is indicated and standard value is summed with one-

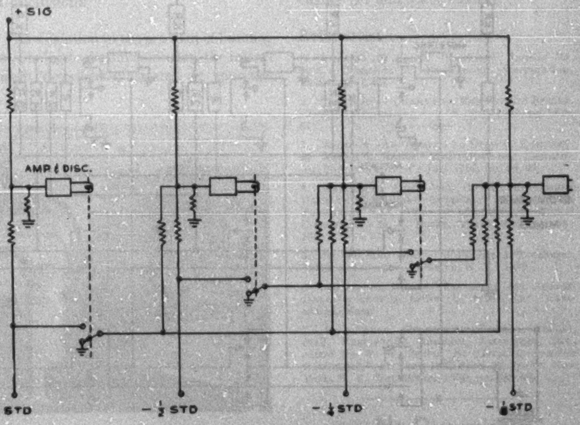


Fig. 3. Analog-to-binary digital converter

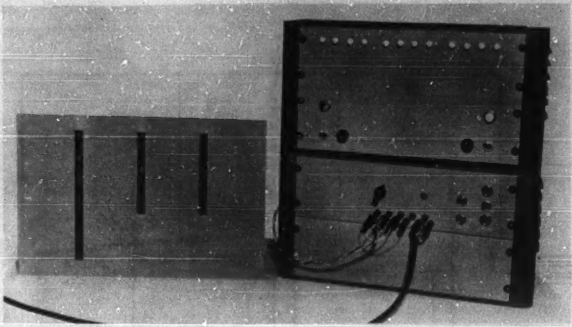


Fig. 4. Quantizer. Left, indicator panel. Right, top to bottom: quantizer proper, power supply, control circuit, and readout

half standard in the next stage of comparison, whereas an unknown value smaller than standard causes a zero to be indicated and zero value to be summed with one-half standard. The unknown value is again compared with the above sum. A signal greater than this sum again causes a binary one to be displayed while one-half standard is summed with one-quarter standard plus the zero or unity obtained from the first operation. Alternately a signal value smaller than this sum, resulting in a recorded zero, causes zero value to be summed with the one-quarter plus zero or unity obtained from the first operation.

This process, theoretically, can be continued for as many binary digits as needed. As shown in Fig. 3, signal voltage is summed against a minus standard in the first stage. A polarity (phase) discriminator establishes the appropriate binary digit and operates a switch accordingly, switching zero or standard voltage to the next summing circuits. The successive phase or polarity discriminators operate their switches accordingly. When these operations go to conclusion, the switch positions indicate the binary reading.

Model Constructed

To more readily construct a progressive code quantizer for demonstration purposes, and to minimize the number of vacuum tubes, a quantizer was constructed which effectively accomplishes the same purpose as reversing the output phase of the fixed-gain amplifiers of Fig. 2 by reversing everything else instead. This was done in a way that eliminated the need for all the fixed-gain amplifiers, with their cumulative distortion and phase shift. This quantizer is shown in Fig. 4.

QUANTIZER PROPER

The circuitry for the first four stages of the quantizer is shown in Fig. 5. The operation of a phase-discriminator plate relay reverses the plate-voltage phase of all following discriminator plates. It also exchanges the ground potential connection with the standard voltage connection for all following stages. The four stages as shown are able to set up all 16 possible code combinations, the circuit being the same as for the hundreds digit of the quantizer shown in Fig. 4. For the tens and units coded digit, only 10 combinations of the possible 16 are used. The circuitry is modified to achieve this purpose. The 10 combinations chosen are centrally picked within the 16. The resulting symmetry allows some economy of contacts in the readout circuit.

It will be noted that the summing network of Fig. 5 is similar to the summing network of Fig. 3. By having the plate relay of each discriminator stage reverse the a-c plate supply phase of all succeeding stages, and by exchanging ground potential with standard voltage potential for all succeeding stages, an effect is produced which is similar to that produced by the circuit in Fig. 2.

READOUT

The readout circuit for all three decimal digits is as shown in Fig. 6. It will

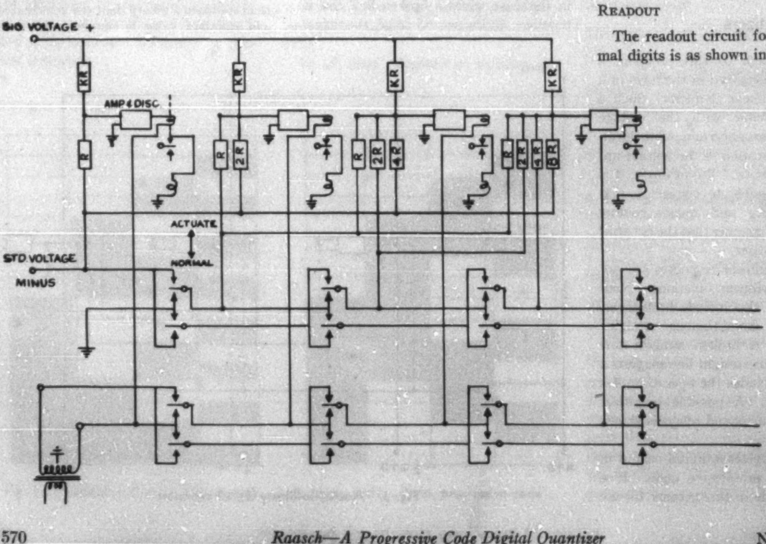


Fig. 5. Hundreds digit circuitry of quantizer shown in Fig. 4

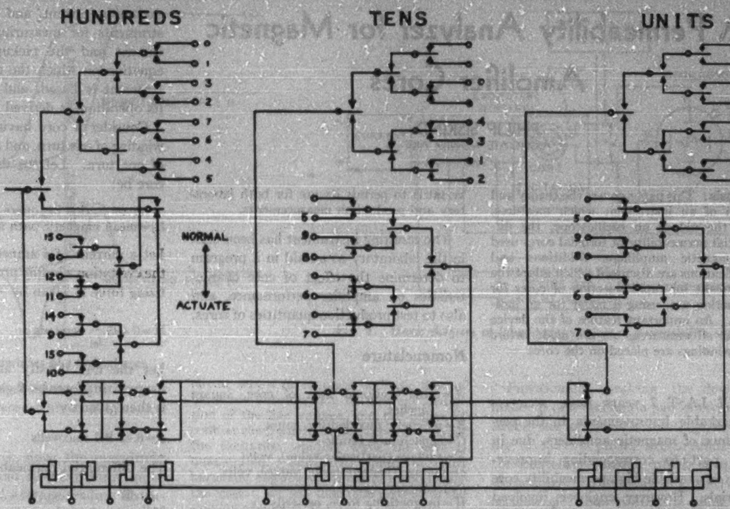


Fig. 6. Readout circuit for translating progressive coded decimal to conventional decimal indication

be noted that a choice between a number and its nine's complement for the tens and units digit output is not only determined by the highest-order Gray-code digit for that coded decimal digit, but also by the reversing contacts, which determine whether the sum of the higher order progressive decimal digits is odd or even. The operation of a single relay in the hundreds digit simultaneously affects the indication of all three decimal digits.

Holding Circuits

The quantizer as constructed also contains circuitry to hold a reading in the readout circuit to allow time for an IBM punch to function. Figure 7 shows the quantizer used in connection with a graph reader and an IBM punch. The graph reader on the left produces an analog voltage, proportional to a graph reading. When an IBM cycle is initiated, the holding circuit holds a reading in the readout. During this time the quantizer proper still follows a varying voltage in the normal manner, and the IBM reproducing punch on the right punches a card with the reading held in the readout. When a release signal is given at the end of the IBM cycle, the readout immediately assumes the setting then indicated by the quantizer.

Conclusions

As a solution to the problem of building an analog-to-digital converter, the instrument described illustrates an approach for which there are many variations. It provides an interesting study in the selection of a numbering system and illustrates the concept of nonlinear elements as the basis for a digital quantizer. It shows that a number system of any radix can be so arranged that any two numbers, only one unit apart, can be indicated by a change of only one digit. It further illustrates that, if the individual digits are indicated by a number of lower radix, the overall number system can still be such that value indications in sequence change progressively with a single digit change per unit value change.

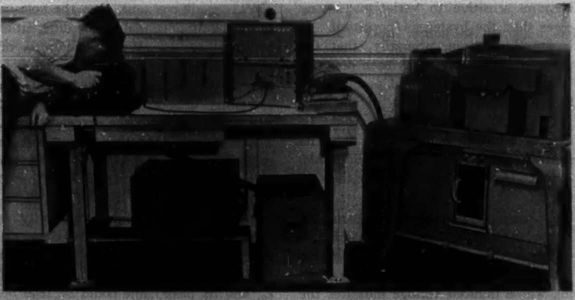


Fig. 7. Graph reader to punch card system

References

1. Type 33-102 SADIC (brochure). *Bulletin No. 3002,* Consolidated Engineering Corporation, Pasadena, Calif.

2. The Binary Quantizer, **Kay Howard Barney.** *Electrical Engineering,* vol. 68, Nov. 1949, pp. 962–67.

3. Anodige—An Analog-to-Digital Converter, **Milton L. Kuder.** *Radio-Electronic Engineering,* New York, N. Y., July 1952, pp. 20, 27.

4. Review of Input and Output Equipment Used in Computing Systems. *AIEE Special Publication S-53,* "Survey of Analog-to-Digital Converters," **H. E. Burke,** March 1953, pp. 98–105.

5. Nonpulsing Digital Converter (brochure). Taller and Cooper, Brooklyn, N. Y., 1952.

6. Automatic Digital Recorder for Analog Data (brochure). Arthur D. Little, Inc., Cambridge, Mass.

7. High-Speed Computing Devices (book), **Staff, Engineering Research Associates, Inc.,** edited by **W. W. Stifler, Jr.** "Data Conversion Equipment," McGraw-Hill Book Company, New York, N. Y., first edition, 1950, pp. 385–93.

◆

No Discussion

A Permeability Analyzer for Magnetic Amplifier Cores

PHILIP SISKIND
ASSOCIATE MEMBER AIEE

Synopsis: This paper covers the theory and design of an instrument which measures, with the aid of an oscilloscope, the differential permeability of toroidal cores used in magnetic amplifiers. Additions and modifications are discussed which adapt the instrument for rapid selection of cores for production use, using a meter for an indicator. An important feature of the device is that all measurements are made before wire windings are placed on the cores.

THE LAST 7 years have seen remarkable improvements in the performance of magnetic amplifiers, due in great part to corresponding improvements in available high-permeability core materials. However, engineers involved with the manufacture of amplifiers and those concerned with producing cores are still faced with related problems arising from the inability of practical and reliable amplifier circuits to tolerate the wide variations in the magnetic characteristics of mass produced cores. In time these problems will be solved by the development of more tolerant circuits and further improved techniques of core manufacture. The interim solution, however, is the proper grading and matching of cores to control amplifier uniformity and quality.

This paper describes an instrument for measuring core characteristics, which was developed to fill the requirements of both the long range and immediate solutions. The development was carried out with the following aims in view:

1. The instrument should be capable of accurately measuring core characteristics which are theoretically correlatable with magnetic amplifier characteristics.

2. Since placing test windings on gapless, toroidal cores is a turn-by-turn process, it is time consuming and involves undesirable handling. The mechanical design, therefore, should not necessitate hand winding.

3. The instrument should be sufficiently versatile to permit its use for both laboratory and production measurements.

The resulting instrument has been used in the laboratory as an aid in a program to determine the effect of core characteristics on amplifier performance, and also to test production quantities of cores.

Nomenclature

A = cross sectional area of core, square inches
$\mathbf{B}$ = magnetic flux density, gauss
C = capacitance, farads
E = voltage (particular value), volts
e = voltage (instantaneous value), volts
f = frequency, cycles per second
$\mathbf{H}$ = magnetizing force, oersteds
I = current (particular value), amperes
i = current (instantaneous value), amperes
L = inductance, henries
l_m = mean magnetic path of core, inches
R = resistance, ohms
t = time, seconds
τ = time constant, seconds
μ_d = differential permeability
ϕ = magnetic flux, maxwells

Laboratory Version of Instrument

It has been shown by Dornhoefer[1] that the transfer characteristic (output versus input) of a magnetic amplifier approximates, in shape, the hysteresis loop of its core material. This plot of core flux versus magnetizing force is related to another characteristic, the differential permeability loop, which is a plot of its point slope, or differential permeability versus magnetizing force, see Fig. 1. Since it is a derivative, this characteristic is more sensitive to variations among cores than the hysteresis loop itself. It is for this reason, primarily, that the permeability analyzer has been designed to measure the differential permeability loop. Although the slope of the hysteresis loop, and thus its shape, may be determined from permeability data, its saturation flux may not. It is desirable, then, to provide some means for measuring this value.

A simplified diagram of the instrument as used in the laboratory is shown in Fig. 2. It consists of a single-turn winding to excite the core, a single-turn pickup winding whose induced voltage indicates the flux conditions in the core, a source of excitation current, and the required instruments for measuring the excitation current and the pickup voltage. The equation on which the design of the instrument is based, and which describes its operation, is derived in the following.

Consider a core having an excitation winding of one turn, and a pickup winding of one turn. Let the dimensions of the core be:

a = cross section in square inches
l_m = mean magnetic path in inches

Let a current of i amperes flow through the excitation winding producing a magnetizing force $\mathbf{H}$ given by

$$\mathbf{H} = 0.495\frac{i}{l_m} \text{ oersteds} \tag{1}$$

Let the flux density arising from this magnetizing force be $\mathbf{B}$ gauss. The flux ϕ is then given by

$$\phi = 6.45\,\mathbf{B}A \text{ maxwells} \tag{2}$$

The differential permeability is given by

$$\mu_d = \frac{d\mathbf{B}}{d\mathbf{H}} \tag{3}$$

Multiplying both sides of equation 3 by the quantity $A(dH/dt)\times10^{-8}$

$$\left[A\frac{d\mathbf{H}}{dt}\times10^{-8}\right]\mu_d = \frac{d\mathbf{B}A}{dt}\times10^{-8} \tag{4}$$

If values obtained for $\mathbf{H}$ and $\mathbf{B}$ from equations 1 and 2 are substituted in equation 4

$$\left[\frac{3.2A}{l_m}\times10^{-8}\frac{di}{dt}\right]\mu_d = \frac{d\phi}{dt}\times10^{-8} \tag{5}$$

The right hand side of this equation may be recognized as the voltage induced in the pickup winding by the changing flux ϕ. Calling this voltage e_μ

$$e_\mu = \left[3.2\frac{A}{l_m}\frac{di}{dt}\times10^{-8}\right]\mu_d \tag{6}$$

If the time rate of change of current di/dt is made constant, then the voltage e_μ will be proportional to the differential permeability μ_d. When the rate of change is positive, the differential permeability will be associated with the ascending arc of the hysteresis loop; when negative, with the descending arc. The standard magnetizing force will be considered as that produced by such a time-linear current. Then equation 6 will describe the operation of the instrument. The differential permeability is evaluated by measuring e_μ; the magnetizing force by measuring i. The flux in the core is proportional to the integral of e_μ, and its peak value could be read on an average-reading instrument. Unfor-

---

Paper 53-290, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 18, 1953; made available for printing May 15, 1953.

Philip Siskind is with the Sperry Gyroscope Company, Great Neck, N. Y.

This paper is a condensed version of the work submitted by the author in partial fulfillment of the requirements for a Master's Degree at Polytechnic Institute of Brooklyn, N. Y.

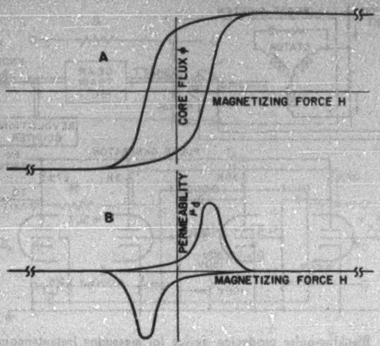


Fig. 1. Characteristics of a typical core

A. Hysteresis loop
B. Permeability loop

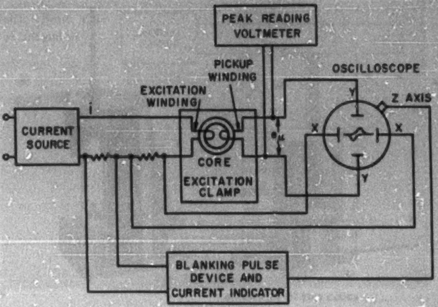


Fig. 2. Block diagram of laboratory, version of permeability analyzer

tunately, however, since the magnitude of e_μ is small and its wave form quite peaked, standard average-reading instruments are incapable of indicating its average accurately. To solve this problem of instrumentation, the voltage is integrated by an electronic integrator, whose peak output voltage is read on a peak-reading vacuum-tube voltmeter.

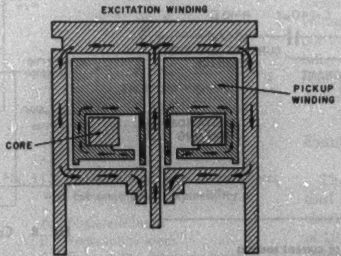
The necessary components of the permeability analyzer are:

1. The excitation clamp—This device is a mechanical arrangement for placing an excitation winding and pickup winding on the core. In order to eliminate whatever effects may be due to dissymmetry of the excitation field, the windings have been designed to produce a standard uniform field, whose direction is always tangential to the core. The windings, therefore, are made up of cylindrically symmetrical shells, nesting one inside the other, completely linking the core and coaxial with it. A cross section of the configuration showing the position of the core, the path of the excitation current, and the direction of the induced voltage e_μ appears in Fig. 3. To permit insertion of the core, both windings are made in separable parts, as shown in Fig. 4. The windings are mounted so that the cap (upper part) of the excitation winding is raised by a rack and pinion operated by a crank. The pickup winding is lifted by hand to insert the core, see Fig. 5.

2. The current source—The current is derived from a sinusoidal line voltage through a high series impedance, as diagramed in Fig. 6. The time linearity requirement is met with reasonable accuracy only at angular arguments near zero. The effect of nonlinearity at large angles is made negligible by adjusting the peak magnitude of current to insure virtual saturation of the core within a few degrees of the zero-axis crossing. The series impedance consists of a series inductance-capacitance combination, resonant at the supply frequency, and a resistor to limit the current. The resonant circuit serves to suppress both the flow of harmonic currents due to inherent distortion of the line voltage, and the pulses of counter electromotive force appearing across the excitation winding while the core is unsaturated. Calculated values of third harmonic distortion, which is theoretically the largest, are less than 0.05 percent in the existing system.

3. Indicating instruments—The quantities e_μ and i are measured with the aid of an oscilloscope. The voltage e_μ is applied to the vertical deflection system, while the voltage drop across a metering resistor, through which the current flows, is applied to the horizontal deflection system. The resulting pattern on the screen will then be a differential permeability loop similar to that shown in Fig. 1. It is difficult to record this pattern with reasonable accuracy when the horizontal gain is increased to magnify the portion of interest. Magnification causes the current peak, and often the current zero, to be thrown off the screen, leaving no simple means of determining current values. Photographic recording is unsatisfactory for the same reason, and also because comparison of a large number of curves is inconvenient. To eliminate this difficulty, the screen is calibrated only in the vertical direction, by simultaneously reading the peak value of e_μ on a vacuum-tube voltmeter, and using a device (described in the following) to transcribe the permeability loops point by point.

Fig. 3. Cross section of excitation and pickup windings

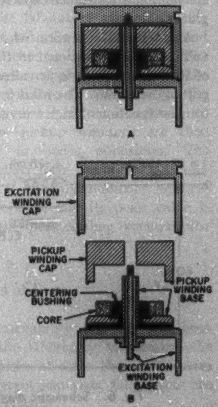


Functionally speaking, the device is made up of a dial and a pair of terminals. Across the terminals is a voltage, consisting of narrow pulses synchronous with the excitation current, which is applied to the z-axis (cathode-ray tube grid) of the oscilloscope, thus causing a blank spot to appear in the trace. The dial reading is proportional to the instantaneous value of current when the pulses occur. The trace is transcribed by recording the vertical deflection of the spot and the corresponding dial reading.

Functions of the device are performed by the equipment shown in the diagram in Fig. 7. The blanking pulse is produced by a pulse generator, which is triggered when its sinusoidal input voltage passes positively through zero. This

Fig. 4. Construction of excitation and pickup windings

A. Closed position
B. Open position



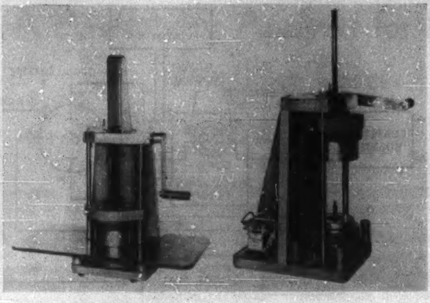


Fig. 5. Laboratory model of permeability analyzer

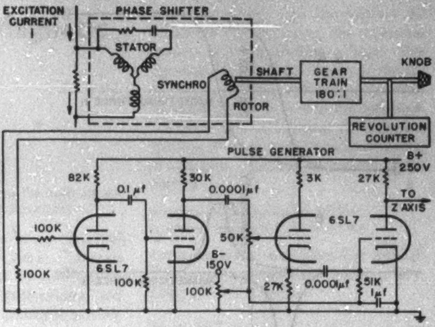

EXCITATION CURRENT
I
PHASE SHIFTER
STATOR
SYNCHRO
ROTOR
SHAFT
GEAR TRAIN 180:1
KNOB
REVOLUTION COUNTER
PULSE GENERATOR
B+ 250V
82K
0.1 μf
30K
0.0001μf
3K
27K
TO Z AXIS
100K
6SL7
50K
B- 150V
100K
6SL7
100K
100K
27K
0.0001μf
51K
1μf

Fig. 7. Blanking-pulse producing device for measuring instantaneous current

input is a voltage shifted in phase from the excitation current by an electromechanical phase shifter whose shaft rotation is equal in radians to the phase shift. Thus, the angle at which the pulse occurs is known, and the instantaneous value of current at the time of the pulse is proportional to the sine of the angle. Values of current in the linear range are considered proportional to the angle itself.

The pulse generator is of conventional design; the phase shifter is a synchro whose stator windings are supplied with 3-phase voltages derived from the excitation current by a phase-splitting network. The rotor shaft is rotated by a knob coupled to it through a stepped-down gear train.

The integrator—a diagram of the integrator used to determine core flux is shown in Fig. 8. A brief analysis of the circuit will show its superiority over the conventional resistance-capacitance integrator.

A 4-terminal network whose output voltage e_o is proportional to the integral of its input voltage e_i is called an integrating network. The ideal form of the transfer function of such a network is

$$\frac{e_o}{e_i}=\frac{1}{\tau}\left[\frac{1}{p}\right] \tag{7}$$

where $1/\tau$ is the constant of proportionality. Practically, this function can only be approximated by one of the form

$$\frac{e_o}{e_i}=\frac{1}{\tau}\left[\frac{1}{p+\frac{1}{\tau}}\right] \tag{8}$$

The approximation improves indefinitely as τ increases, but at the same time the constant of proportionality decreases, thus reducing the magnitude of the output voltage.

Two networks having transfer functions of the form of equation 8 are shown in Fig. 9. The first is a conventional resistance-capacitance integrator whose transfer function is

$$\frac{e_o}{e_i}=\frac{1}{RC}\left[\frac{1}{p+\frac{1}{RC}}\right] \tag{9}$$

Here $\tau=RC$. The second network is a Miller-effect integrator which involves the use of an amplifier gain, G, having capacitive voltage feedback. The transfer function is

$$\frac{e_o}{e_i}=-G\left[\frac{\frac{1}{(1+G)RC}}{p+\frac{1}{(1+G)RC}}\right] \tag{10}$$

Here, the value of τ for integration purposes is $(1+G)RC$, but the constant of proportionality is

$$\frac{-G}{(1+G)RC}\approx-\frac{1}{RC}, \text{ when } G \text{ is very large} \tag{11}$$

Thus for equal performance as an integrating network, the Miller integrator has an output approximately G times that of the resistance-capacitance integrator. The effect is the same as adding an amplifier of gain G to a conventional integrator; however, variations of gain and the effects of amplifier phase shifts will be considerably reduced, as long as the gain is high enough.

The amplifier used in this equipment consists of a "starved pentode" (operated with a very high plate load) feeding a

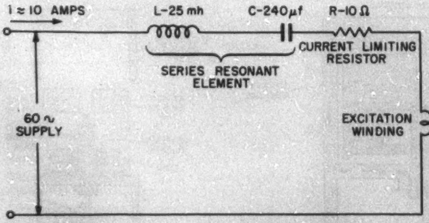

I ≈ 10 AMPS
L-25 mh
C-240μf
R-10Ω
SERIES RESONANT ELEMENT
CURRENT LIMITING RESISTOR
60 ∿ SUPPLY
EXCITATION WINDING

Fig. 6. Schematic diagram of current sources

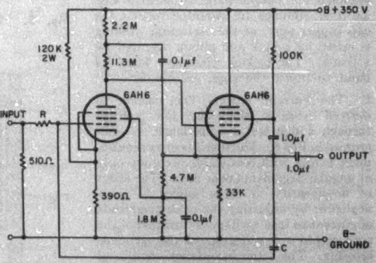

B + 350 V
2.2 M
120K 2W
11.3 M
0.1μf
100K
6AH6
6AH6
INPUT
R
1.0μf
OUTPUT
1.0μf
510Ω
4.7 M
390Ω
33K
1.8 M
0.1μf
B- GROUND
C

Fig. 8. Circuit used with permeability analyzer for flux measurement

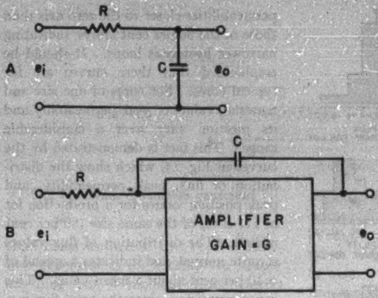


Fig. 9 (left). Basic diagrams of integrators

A. Conventional integrator
B. Miller-effect integrator

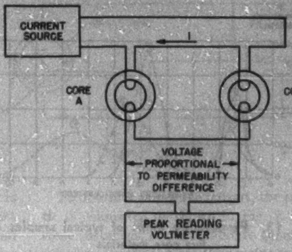


Fig. 12. Arrangement for matching the characteristics of two cores

cathode follower. Since there is only one phase inversion, the performance is similar to that of a 1-stage amplifier, and oscillation problems are considerably reduced. The gain of the amplifier is about 2,500 and its cutoff frequency 1 kc. The upper frequency limit of the integrator for 1 per cent accuracy is about 16 kc, which is about the 250th harmonic of a 60-cycle fundamental. The time constant for integration purposes is about 6.5 seconds, when the resistance-capacitance product is adjusted to cause zero attenuation at 60 cycles.

Core Testing in Production

The most stringent limitation on the variation of cores used in quantity production of amplifiers is that placed on the groups of two or four cores forming the complement of an individual amplifier. The transfer curves of the component reactors must be sufficiently alike to prevent deterioration of performance due to unbalance. It is desirable that this alikeness be achieved by using reactors with matched cores, rather than by separate bias adjustments on each reactor, since the latter procedure involves an increase in the number of components and a decrease in the reliability of the amplifier.

The principal effect of differences of peak permeability and saturation flux between cores will be differences of slope of the transfer characteristic. Far more influential is the position of the permeability peak, which causes an equivalent biasing effect on the transfer characteristic. This effect is illustrated by the curves in Fig. 10, showing the transfer characteristics and permeability loops for two cores. It will be noticed that the displacement of the transfer curves in the *H* direction corresponds in magnitude, to the displacement of the peak permeabilities.

In production, sets of two or four cores, whose characteristics are matched, are selected from run-of-the-mill stock by a 2-step process. First, the stock is divided into grades, within which variation of flux and peak permeability values is small. The second step in the selection consists of grouping the cores in each grade into the required matched sets. It is necessary, for this step, to have a rapid means of detecting differences in peak position. Using the laboratory version of the instrument to perform this function would result in a slow and relatively tedious process. This objection may be overcome by using the following technique.

A voltage is developed which is proportional to the instantaneous difference in permeability between two cores. Since cores in one grade have a small difference in peak permeability, the magnitude of this voltage will be due primarily to the displacement between the permeability loops. This is demonstrated by the curves in Fig. 11, which show the hysteresis and the permeability loops for two cores whose peak permeabilities are the same but whose loops are displaced from each other. The difference between the loops due to this displacement is plotted on the same graph and will be seen to be of considerable magnitude. Thus, limiting the values of the "difference voltage" is a sensitive means of limiting displacement. Development of this voltage involves the same equipment used in the laboratory version of the instrument, with the addition of a second excitation clamp; see Fig. 12. The two cores are compared by connecting the excitation windings in series so that the same magnetizing force is applied to each. The pickup windings are then connected in series opposition. The voltage across their combination is proportional to the instantaneous permeability difference, and its peak value is read on the peak-reading vacuum-tube voltmeter.

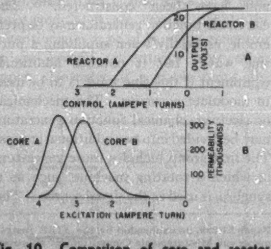


Fig. 10. Comparison of core and reactor characteristics for two cores of the same material

A. Transfer characteristics
B. Permeability loops

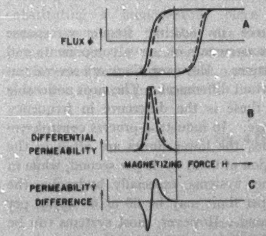


Fig. 11. Comparison of loops for two cores with the same peak permeability

A. Hysteresis loops
B. Permeability loops
C. Permeability difference curve

Examples of Data

The permeability analyzer has been used in the laboratory to measure the characteristics of more than 1,000 cores of 12 different materials, falling into the following classes:

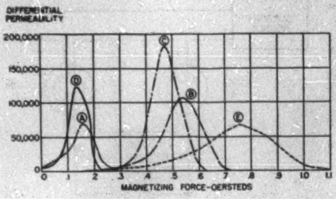


Fig. 13. **Permeability loops of typical samples of five core materials**

A. 80 per cent nickel-steel alloy
B. 50 per cent nickel-steel alloy
C. 50 per cent nickel-steel alloy, grain oriented
D. 80 per cent nickel-steel alloy, high purity
E. Silicon-steel alloy, grain oriented

1. 80 per cent nickel alloy (Numetal, Hymu 80, etc.)

2. 50 per cent nickel alloy (Hipernik, 4750, etc.)

3. 50 per cent nickel alloy, grain-oriented (Deltamax, Orthonol, etc.)

4. 80 per cent nickel alloy, highly purified (Supermalloy)

5. Silicon steel, grain-oriented (Hipersil)

The cores are all toroids, wound from thin ribbons, and are of several sizes.

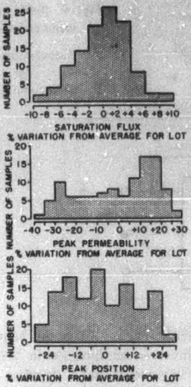


Fig. 14. **Distribution of measured values for a lot of 111 cores**

The curves in Fig. 13 are differential permeability loops for materials in each of the above classes, demonstrating the wide variety of characteristics presently available. (Since the permeability loops are symmetrical, only one half of each is plotted.) Of the nickel alloys, the 80-per-cent alloys will be seen to have peak permeabilities closer to the zero axis than those of the 50 per cent alloys, indicating narrower hysteresis loops. It should be emphasized that these curves are for typical cores. For cores of one size and material, values of peak permeability and its position vary over a considerable range. This fact is demonstrated by the curves in Fig. 14, which show the distribution of flux, peak permeability, and peak position values for a production lot of 111 cores of the same size (80 per cent nickel). The distribution of flux values is quite normal, and indicates a spread of ±10 per cent about a mean value. This is not surprising, since the total flux is a function of the annealing process to only a slight extent, and depends mainly on the analysis of the alloy and the amount of material in the core. On the other hand, the peak permeability and its position are very much dependent on the annealing and show variations which appear to be random over a much larger range. It is this range, especially for values of peak position, which demonstrates clearly the nature of the problem.

Reference

1. Self-Saturation in Magnetic Amplifiers, W. J. Dornhoefer. *AIEE Transactions*, vol. 68, part II, 1949, 835–50.

No Discussion

Forcing Function Generator Employing Conductive Plastic

L. W. NORMAN
ASSOCIATE MEMBER AIEE

Synopsis: This paper describes a forcing-function generator which uses a conductive-plastic potentiometer to provide a forcing-function signal for testing servo systems and components. The unit described gives a sine-wave signal, but methods for generating a square wave or triangular wave are also discussed. A phase-measuring device is made an integral part of the function generator, and allows output versus input phase shift to be read directly from a calibrated dial.

THE FORCING FUNCTION generator is an important instrument in a servo laboratory. Like the signal generator, which is its counterpart in the communications field, it serves as a signal source in making frequency-response measurements on servo components and systems. However, there are several important differences. The most noticeable of these is the difference in frequency range. In industrial-process control systems, the lower limits are often in the region of 0.001 cycle per second, while in other systems, especially hydraulic, the upper limit may be 50 or 60 cycles per second. However, most systems can be adequately tested with a function generator which has a range of 0.05 to 20 cycles per second.

Since the function generator's primary use is in frequency-response measurements, the most important signal it must supply is a sine wave. This signal will usually take one of two forms, a simple sine wave or a sine-wave-modulated a-c carrier.

Forcing function generators can be divided into two general groups, mechanical and electrical. From these two general groups there are an almost unlimited number of variations which can and have been constructed.$^{1-5}$ The electronic function generator may be preferable, especially when supplying a pure sine wave; but it requires additional equipment if this sine wave is to be used to modulate a carrier. The mechanical or electromechanical function generators may be divided into two additional groups. The first group includes those generators in which a rotating machine, such as a synchro, is used as the signal source. The

Paper 53-239, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 23, 1953; made available for printing April 29, 1953.

L. W. Norman is with the Minneapolis-Honeywell Regulator Company, Minneapolis, Minn.

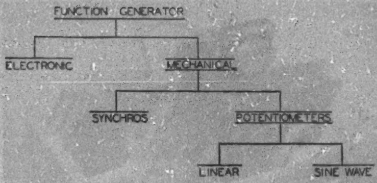


Fig. 1. Types of function generators

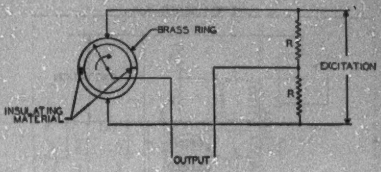


Fig. 3 (right). Method of connecting plastic for the generation of square waves

chief disadvantage of this type of function generator is its inability to supply anything but a modulated a-c carrier signal at low frequencies. With d-c excitation it has the disadvantage of phase shift at higher frequencies. Those of the second group of mechanical function generators employ wire-wound potentiometers and wipers as means of obtaining the desired signals. By applying d-c to the potentiometer, a pure sine wave is obtained; and by applying a-c, a modulated sine wave is obtained. Fig. 1 is a chart showing the various types of forcing function generators. It should be noted that the group of function generators employing potentiometers has been broken down into two additional groups. One group includes those generators which employ a scotch yoke or some similar mechanical device to drive a linear potentiometer. The second group includes those generators employing sine-wave potentiometers and other special devices. The function generator to be described is included in the latter group.

There is also a third group of function generators which might be mentioned. This is a miscellaneous group which employs special alternators, rotating discs, photoelectric cells, and similar devices.[6,7] This group is rather specialized and constitutes but a small percentage of the function generators in general use.

Because it is readily adaptable for supplying either a sine wave or a sine-wave-modulated a-c carrier signal, the potentiometer type of function generator enjoys several advantages over other types. However, it is not without disadvantages. If a linear potentiometer is employed, some mechanical linkage must be employed to produce a sinusoidal signal wave. Mechanical linkages are subject to wear. To a certain extent this objection can be overcome by the use of sinusoidal potentiometers. A second objection, not so easily overcome, is inherent in the potentiometers themselves.[8] As the wiper moves along the potentiometer, its output is a series of successive steps as the wiper moves from one wire to the next. Thus the output signal is not a smooth curve but is actually a series of steps. By making these steps very small, the approximation to a smooth curve becomes much better, but the cost of the potentiometer increases and its life decreases.

Function Generator

The mechanical function generator employing some type of linear potentiometer or sinusoidal potentiometer seems to be the most desirable from a standpoint of reliability and simplicity. The sinusoidal potentiometer is ideal except for the problem of resolution and the possibility of excessive wear. One potentially attractive solution to these problems is the use of a conductive-plastic potentiometer.

It is well known that a wire-wound potentiometer made in the form of a square card can be used to produce sine wave electrical signals with mechanical rotary motion input.[9] However, the problem of resolution still exists. By substituting a homogeneous piece of conductive plastic for the wire-wound potentiometer, the number of steps can be increased to infinity so that a stepless sine-wave signal is produced.

To obtain a sine-wave output, the plastic is connected as shown in Fig. 2. In some instances, it is desirable to generate a forcing function other than a sine wave. This will usually be a square wave or triangular wave. A square wave can be generated quite easily without the use of conductive plastic. Fig. 3 shows a circuit arrangement which might be used for the generation of a square wave. However, conductive plastic can be employed to generate a triangular wave. Fig. 4 shows the circuit arrangement which is necessary.

The type of output which is obtained from the function generator depends upon the type of excitation applied to the conductive plastic. With d-c excitation the output is a sine wave having a frequency equal to the drive shaft speed. When a-c excitation is used, the output becomes a sine-wave-modulated a-c carrier. Here the carrier frequency is the frequency of the excitation and the modulating frequency the drive shaft speed.

The circuits of Figs. 2, 3, and 4 make it necessary to provide some form of external summing circuit if the function generator is to be used in making closed-loop studies. To simplify the testing of both open and closed-loop systems, the above circuits were modified in order that closed-loop systems could be studied without the need for external summing circuits. The circuit of Fig. 5 shows the changes necessary to accomplish this. With the switch in the open position, the function generator is supplying an input signal to the servo system. This signal is independent of the output position of the servo system and gives an open-loop test. When the switch is in the closed

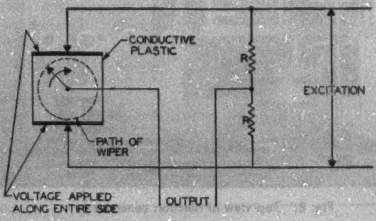
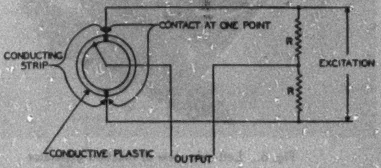


Fig. 2 (left). Method of connecting plastic for the generation of sine waves

Fig. 4 (below). Method of connecting plastic for the generation of triangular waves



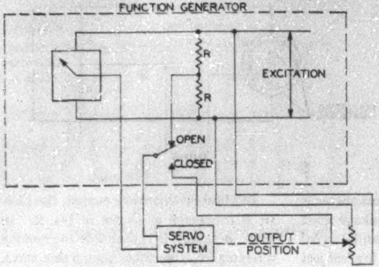


Fig. 5. Circuit for making open- and closed-loop studies

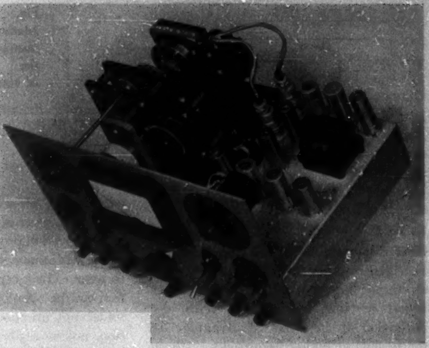


Fig. 7. Right-side view of function generator

position, the signal supplied to the servo system is the output of the function-generator series, summed with a signal proportional to the output position of the system. This gives a closed-loop test.

A 2-phase servomotor with built-in velocity generator, connected in a velocity servo system, was used as the driving motor for a function generator of the type described above. The particular motor used required approximately 7 watts per phase. The maximum speed of the gear-train output shaft was 200 rpm, or 3.7 cycles per second. By suitable adjustment of input voltages to the amplifier, the output speed could be reduced to 3 rpm, or 0.05 cycle per second. To obtain higher speeds, it was necessary to provide a second output shaft in the gear train. This shaft, which was brought out as an extension on one of the existing gears, gave an output speed range of 0.9 to 60 cycles per second. A lever was provided to shift an idler gear between the function generator proper and either of the two output shafts. Manual motion of this lever allowed selection of one or the other of the two speed ranges. Figs. 6, 7, and 8 show the mechanical details of the drive mechanism.

To help eliminate, as nearly as possible, all additional equipment needed, a-c and d-c excitation for the conductive plastic potentiometer was provided within the function generator. A frequency meter was also included to enable quick determination of the forcing frequency. Frequency was determined by rectifying the output voltage of the velocity generator and feeding this rectified voltage into a d-c milliammeter. Fig. 9 gives the complete schematic diagram of the forcing function generator and Fig. 10 shows the front panel of the instrument.

Phase Measurement*

To measure phase shift, a conventional d-c cathode-ray oscilloscope is used. The output wave, produced by a sinusoidal-potentiometer wiper positioned by the servo output shaft, is applied to the vertical deflection plates; and a sawtooth sweep voltage, taken from a linear potentiometer geared to rotate at twice the forcing-function generator's speed, is applied to the horizontal deflection plates. The resulting display consists of the positive half and the negative half of the output wave superimposed on the screen. The output-phase reference point is at the intersection of the two half waves.

Connected to the rotating shaft of the function generator is a shorting bar which provides a momentary short circuit between two brushes. If the shorting bar on the forcing-function generator is wired so as to short out the output wave momentarily, a marker depression will show on one part of the output wave. This can be made more distinct by adding a battery in series with one of the leads to the shorting bar. The position of this

---

* The method of phase measurement is essentially that contained in an unpublished memo by Frank W. Ainsworth, aeronautical research engineer, Minneapolis-Honeywell Regulator Company, Minneapolis, Minn.

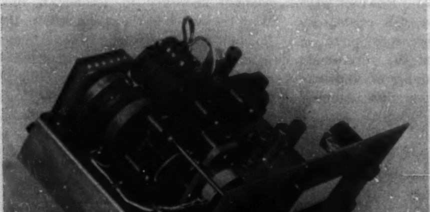


Fig. 6. Left-side view of function generator

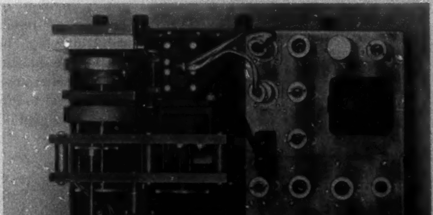


Fig. 8. Top view of function generator

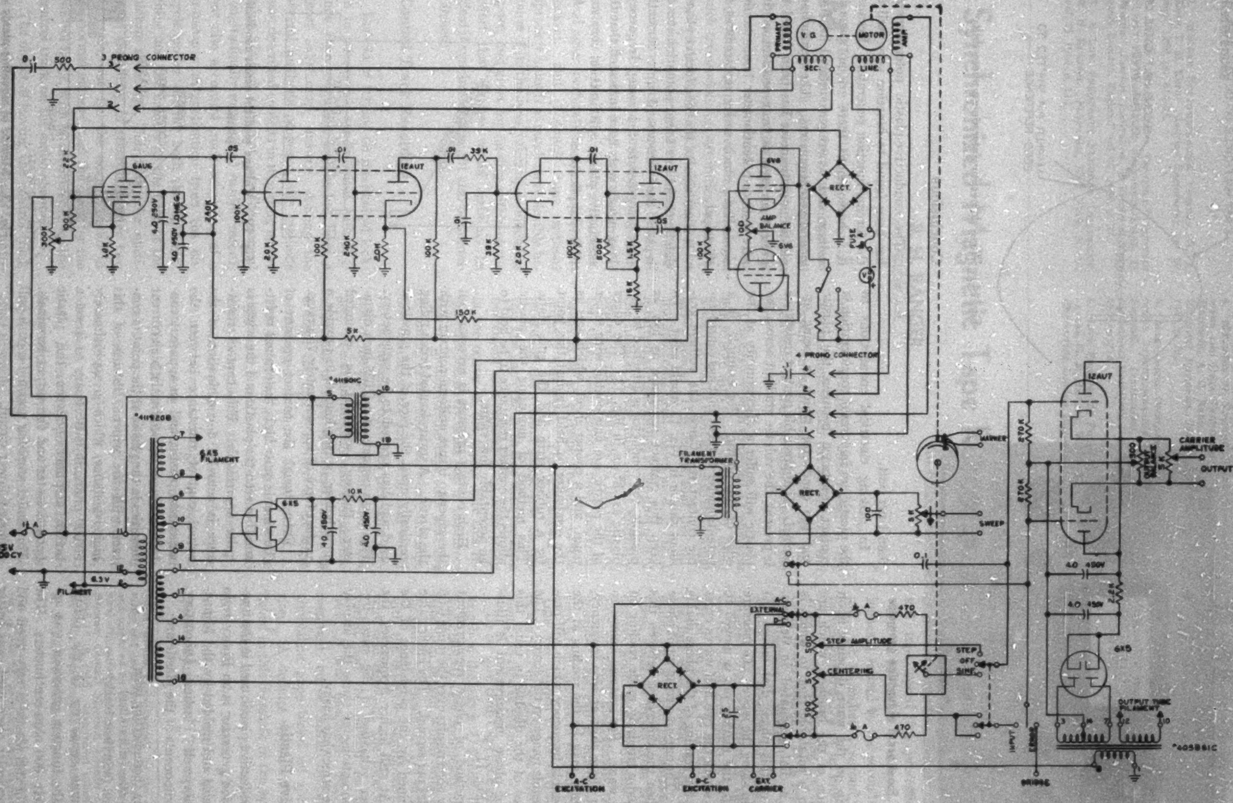

0.1
500
3 PRONG CONNECTOR
6AU6
22K
200K
100K
10K
4.0 250V
4.0 450V
10 MEG
230K
.05
100K
20K
12AU7
.01
240K
20M
.01
35K
38K
150K
5K
500K
1.5K
15K
6V6
AMP BALANCE
100
RECT.
FUSE
V.G.
MOTOR
PRIMARY
SEC.
LINE
AMP
4 PRONG CONNECTOR
*4H501C
*4H920B
6X5 FILAMENT
6X5
10K
4.0 450V
115V 400CY
1 A
6.3V
FILAMENT
FILAMENT TRANSFORMER
SWEEP
MARKER
270K
CARRIER AMPLITUDE
OUTPUT
2.2K
OUTPUT TUBE FILAMENT
*405861C
A-C
EXTERNAL
D-C
STEP AMPLITUDE
CENTERING
470
STEP
OFF
SINE
INPUT
ERROR
BRIDGE
.25
A-C EXCITATION
D-C EXCITATION
EXT. CARRIER

Fig. 9. Schematic of forcing function generator

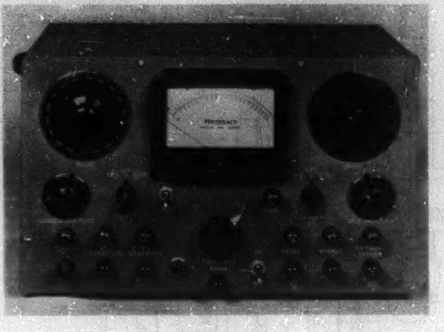


Fig. 10. Front panel of function generator

marker pip on the wave depends upon the position of the brushes, which are carried on a rotatable ring to which is attached a dial graduated in degrees. Rotation of this phase dial moves the marker pip along the wave. To measure phase shift, the marker pip is moved to the point where the two half waves cross over, and the reading of the phase dial is observed. The output-versus-input phase shift is then the difference between this reading and a similarly made reading on the input wave. For systems known to have zero phase shift at zero frequency, the reference may be the reading at the lowest possible frequency which will produce a useable image on the screen. The appearance of the display in using this method of phase measurement is illustrated in Fig. 11.

This method of phase measurement gives very good results over the range 0.05 to 20 cycles per second. The lower value is determined by the persistence of the screen. At still higher frequencies, the high wiper velocities of the sweep-frequency potentiometer cause the display to become somewhat hashy.

Conductive Plastic

The conductive plastic used in the forcing function generator is an Epoxy resin compounded and molded by the Minneapolis-Honeywell Plastics Laboratory. Major components of the plastic are:

1. Epoxy thermosetting resin
2. Inert filler (silica filler)
3. Carbon (acetylene black)

The volume resistivity of the plastic used in the function generator was approximately 900 ohm-centimeters. The linearity of this plastic is very good and in the particular piece used, was better than 1 per cent.

In making electrical connections to the plastic, an important consideration is contact resistance between the plastic and the conductor which is being attached. If appreciable contact resistance is present, the voltage gradient across the plastic will suffer a discontinuity at the edge of the plastic. For example, in a potentiometer element that has considerable contact resistance, the voltage gradient shown in Fig. 12 would exist. To keep this contact resistance as low as possible, the conductor is attached to the plastic as shown in Fig. 13. The edges of the plastic are first coated with silver paint, and a copper sheet, which acts as the conductor, is held against this paint by a rubber strip and the end plate. This method of mounting the plastic allows the contact resistance to be held to an acceptably low value.

High contact pressure is necessary on the wiper to insure uniform contact with the plastic. A spring-loaded brush riding vertically upon the face of the plastic is used and has proved to be satisfactory. This brush consists of a 0.025-inch diameter Paliney wire. It is mounted in a holder, spring-loaded to provide a contact force of approximately 75 grams.

Even though the contact pressure of the wiper is quite high, the contact resistance between the plastic and the wiper is quite appreciable. This contact resistance is actually a combination of the actual contact resistance between the wiper and plastic, and an apparent contact resistance caused by the high current densities in the plastic at the point of contact with the wiper. Because of this high resistance, which is approximately 50,000 ohms, it is necessary to insert a buffer amplifier between the plastic potentiometer and the output terminals. This buffer amplifier uses a d-c push-pull cathode-follower circuit.

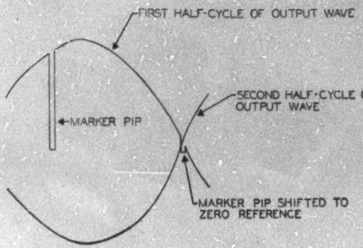

FIRST HALF-CYCLE OF OUTPUT WAVE
SECOND HALF-CYCLE OF OUTPUT WAVE
MARKER PIP
MARKER PIP SHIFTED TO ZERO REFERENCE

Fig. 11. Wave shapes for determining phase shift

Conclusions

The conductive-plastic potentiometer has been found to function very well in the forcing-function generator. Because it is homogenous there are no steps in voltage such as are encountered in wire-wound potentiometers. The conductive-plastic potentiometer has been found to present as great a wear problem as the wire-wound potentiometer, but its design allows much easier replacement of the wiper.

It is anticipated that conductive plastic can be useful in a linear potentiometer to replace the wire-wound unit currently used to generate the sweep frequency. Preliminary tests have indicated that the conductive plastic will function very well in this application.

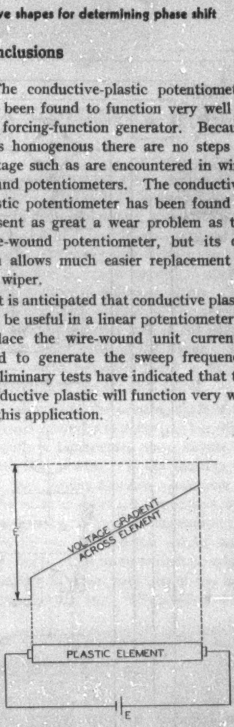


Fig. 12. Voltage gradient across conductive plastic

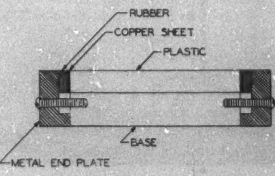


Fig. 13. Method of mounting plastic

References

1. Modern Servomechanism Testers, G. A. Korn, T. M. Korn. *Electrical Engineering*, vol. 69, Sept. 1950, pp. 814–16.

2. Phase Calibrated Variable-Frequency Supply for Testing of Servomechanisms, D. Morris. *Journal*, Institution of Electrical Engineers, London, England, vol. 97, 1950, p. 37.

3. A Low Frequency Function Generator, R. H. Brunner. *Electronics*, New York, N. Y., vol. 25, 1952, p. 114.

4. Methods of Testing Small Servomechanisms and Data Transmission Systems, E. W. Marchant, A. C. Robb. *Journal*, Institution of Electrical Engineers, London, England, vol. 94, 1947, p. 292.

5. Mechanical Oscillators and Their Electrical Synchronization, S. W. Herwold, R. W. Gemmel. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 68, 1946, p. 713.

6. Low Frequency Photo-Mechanical Oscillators, G. R. Baldock, W. G. Walter. *Electronic Engineering*, London, England, vol. 17, 1945, p. 326.

7. A Low Frequency Alternator, E. B. Kurtz, M. J. Larson. *Journal*, Institution of Electrical Engineers, London, England, vol. 27, 1939, p. 148.

8. Catalog. The Helipot Corporation, South Pasadena, Calif.

9. Electronic Instruments (book), D. McRae. McGraw-Hill Book Company, Inc., New York, N. Y., 1948, p. 162.

◆

No Discussion

Synchronized Magnetic Tape Recording

R. H. RANGER
FELLOW AIEE

MOTION PICTURES have obviously required synchronized sound to match picture and words in all released prints. Perforated film has been the usual answer to this requirement ever since sound was added on film. But now that magnetic recording has shown such tremendous opportunities in quality and economy, the use of this facility for sound for motion pictures is immediately attractive. One answer has been, of course, the use of normal photographic film stock, coated however with iron oxide instead of the usual silver emulsions. And again perforations accomplished the lock-in. But 1/4-inch tape without sprocket holes is so much more flexible because of its thinness, takes up so much less space, and gives top quality with so little effort that its use would certainly be ideal if synchronism might also be accomplished.

Control Track Recording

Instead of relying on mechanical lock-in to accomplish synchronism with 1/4-inch tape, electronic lock-in has been a most satisfactory answer. Among the several methods proposed, the methods here described consists in recording on the same 1/4-inch tape not only the normal longitudinally-magnetized audio record, but also a control track consisting of a transverse magnetization from the 60-cycle power driving both the tape machine and the camera at the time of the picture shooting. This makes for maximum simplicity in recording, gives two channels, one for the normal audio in the normal manner, and the other for the control track, with no interference by the control track on the audio channel.

The object of having the control track is to provide means on playback for automatically controlling the speed of the tape machine so that the audio track will remain in synchronism with the film, regardless of the absolute speed of the film.

It is easy enough to record a transverse flux in step with the 60-cycle power, using a head turned at right angles to the normal audio heads. This means that the control recording head shown at *A* in Fig. 1 is turned at right angles to the direction of the motion of the tape and is normally placed to record directly in the center of this tape. Gap width in most magnetic heads is of an extreme fineness—of the order of half a mil. The first answer to recording the control track was to use this fine gap exactly in line with the tape movement and register a 60-cycle track in the center of the tape. But this required a relatively short gap in the direction of motion of the tape such that the recording of the 60 cycles where the tape first contacted the gap in the head would not be cancelled out when this same portion of the tape hit the trailing end of the gap. To avoid this, and also to make the alignment of tape with head much less critical, the gap *T* is tiled at an angle of 6 degrees as shown on *A* in Fig. 1. This means that as the tape moves to the right and a small voltage from the 60 cycles is applied to the coils of head *A*, the various phases of the recording cycle are put down in shingle fashion over each other along the tape as indicated at S. This tilt as selected is downwards from left to right looking at the front of the head with the tape passing over it from left to right. It could, of course, have been exactly the other way just as well. In fact it is possible to place two independent control tracks on a single tape, one tilting downwards and the other upwards.

In order to prevent this shingled magnetic recording track from affecting the audio track, the length of the tilted gap is most important. The primary standard tape speed for professional tape work is 15 inches a second, with $7^1/_2$ inches a second as the secondary standard. At 15 inches a second, the tape moves forward 1/4 of an inch for every 1/60 second representing 1 cycle of a 60-cycle power source. It is seen that at any instant a vertical gap would cut across a succession of these shingled phases. The resolution of the flux F_0, for the 360-degree point is shown at the left of the gap *V* in *B* of Fig. 1, with F_2 the flux which alone would affect the audio gap *V*. But the summation of all these small F_2 fluxes from zero to 360 degrees would add vectorially to zero to give no effect magnetically on gap *B*, provided this gap is cutting only from zero to 360 degrees of the shingled fluxes. This will happen if the effective length of the recording gap in the direction of F_2 along the tape is 1/4 inch.

To be sure that the effective magnetic length of the gap is 1/4 inch, the matter was checked very completely by using a fixed gap length which was accurately measured physically at 1/4 inch. Then, with the 60-cycle control-recording current held constant, the tape speed was varied and the pickup in the audio head *B* was measured. Fig. 2 shows resultant curve made by calculating equivalent gaps based on speed variations. A definite minimum was established in this crosstalk in the audio channel which was down at least 10 decibles from what the normal 15-inch speed would have given without this accurate speed adjustment. The speed turned

---

Paper 53-300, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 23, 1953; made available for printing May 18, 1953.

R. H. Ranger is with Rangertone, Inc., Newark, N. J.

The author acknowledges with thanks the collaboration of Paul M. Brubaker, chief engineer, Rangertone, Inc.

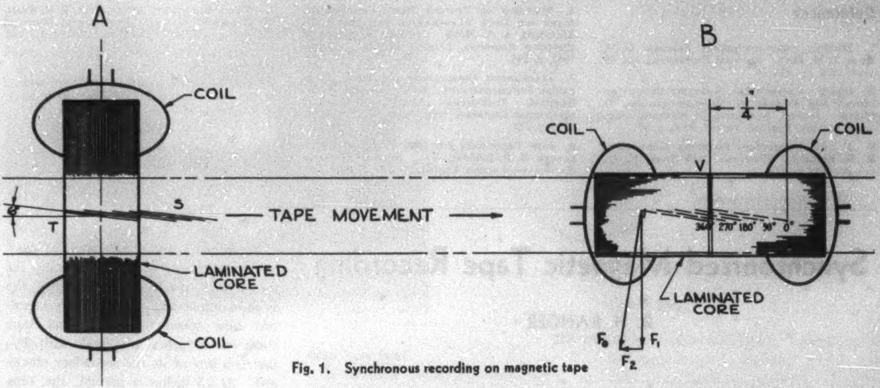


Fig. 1. Synchronous recording on magnetic tape

out to be 15.45 inches per second for this minimum reading. This showed that with the physical length of the gap of 0.250 inch the actual recording was spreading out to be 15.45/60 = 0.2575 inch.

By interpolating from this, it was determined that the actual gap length along the tape should be 0.00242 inch instead of 0.0025 inch, for minimum crosstalk at the regular 15-inch-per-second speed.

With an impedance of approximately 200 ohms at 60 cycles on the recording head, 0.25 volt gives an ample record of the 60-cycle pulse on the tape; 150 microvolts are easily obtained directly off the control playback head. On the audio pickup head crosstalk from this control track will be of the order of 53 decibels down from the normal audio signal.

The synchronizing playback-control head is made very similarly to the synchronizing record head, except that the angle is increased to 10 degrees, which merely means that the tape can have greater excursion up and down. The gap width for control recording is 2 mils and for playback is 9 mils.

Tape Path

In order to provide a very smooth controllable tape path, which will have a minimum of flutter, the tight loop drive has been initiated on 1/4 tape on this equipment, as shown in Fig. 3 for the top plate. The tight loop is the small section in the center from the drive capstan in the very center down to the flywheel at the bottom. This tight loop has been used for motion picture film for several years, except that with tape there can be no sprocket wheels to maintain the loop. The rubber idlers hold the tape in contact with the capstan on each side, and the rotation of this casptan communicates itself to the tape and the tight loop down to the flywheel and back. It would seem that the amount of tape going in would be exactly equal to that going out so that there would be no tension on tape in the tight loop. But it is surprising to see that the tape in the tight loop actually tightens up even with only the single diameter on the capstan. This is because there is an actual tension balance in the tape going in and going out, determined practically entirely by the tension on the tape accomplished by the drag of the supply reel in the upper left of the top plate.

This tension is accomplished by applying the right amount of direct current to the induction motor connected to the supply reel. This makes the tape tension at the capstan, where it enters on the left, greater than that on the right, where the tape leaves. This tension means that the tape will actually be stretched slightly at entering more than it is on leaving. Therefore, it takes more rotation of the capstan to put tape into the loop than it takes to pull it out. The net result is that the amount of tape in the loop decreases up to the time when a balance point is reached when the tension on the inside of the loop equals that of the tape entering from the supply reel.

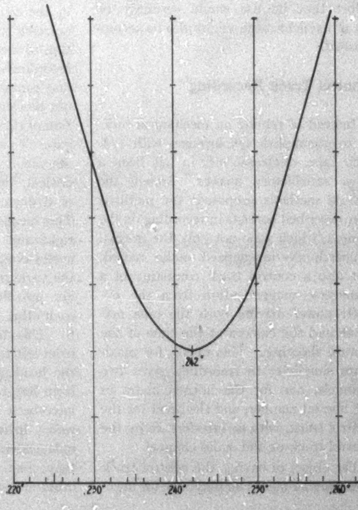


Fig. 2. Crosstalk as a function of recording gap length

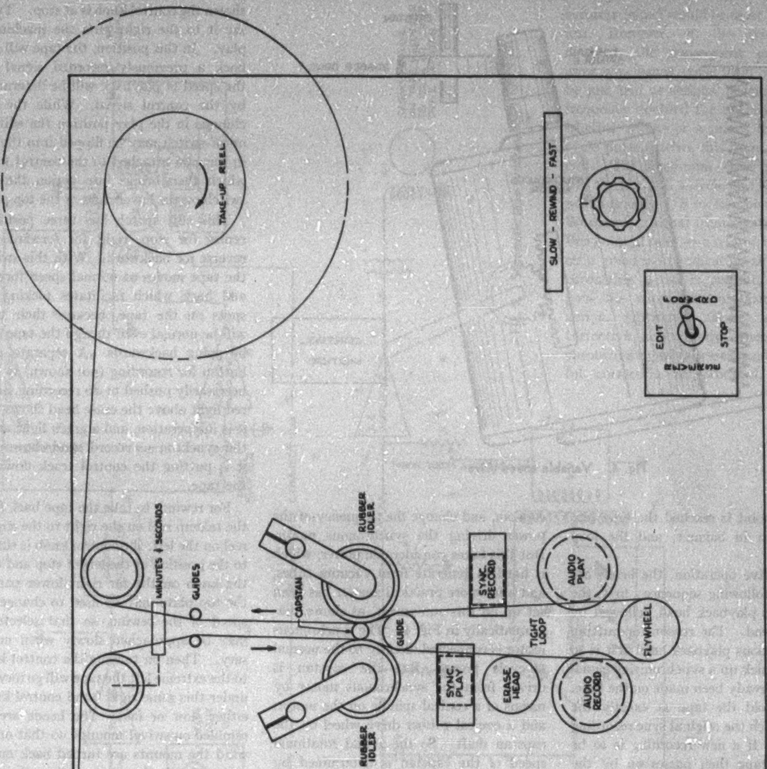

TAKE-UP REEL
SLOW - REWIND - FAST
EDIT
FORWARD
REVERSE
STOP
RUBBER IDLER
MINUTES & SECONDS
GUIDES
CAPSTAN
GUIDE
SYNC. RECORD
AUDIO PLAY
TIGHT LOOP
FLYWHEEL
SYNC. PLAY
ERASE HEAD
AUDIO RECORD
RUBBER IDLER
FAST
SLOW
SPEED CONTROL
7½"
15"
LO
HI
TAPE SPEED
PLAY
STOP
REWIND
FORWARD
OFF
BRAKE
ON
ON
EDIT
OFF
SUPPLY REEL

Fig. 3. Top plate showing tape path

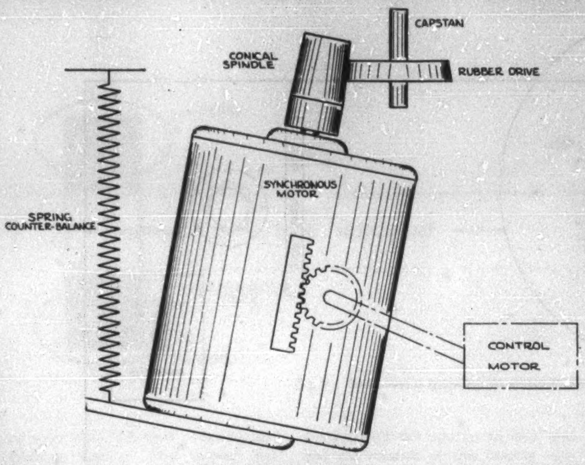


Fig. 4. Variable speed drive

After this point is reached the tape tensions remain in balance, and the loop stays tight.

For effective operation, the heads are put in the following sequence: first, the synchronous playback head, followed by the erase head. The reason for putting the synchronous playback head first is so that it will pick up a synchronizing signal if one has already been made on the tape. This will hold the tape at exactly the speed in which the original sync recording was made. If a new recording is to be made, the tape then passes on by the erase head. The erase head has a transverse wide gap supplied from a high frequency source: 80,000 cycles is currently used. This completely erases audio and synchronizing signals from a tape that is already recorded. The tape then passes on to the audio record head, which has a very narrow gap of the order of half a mil in thickness, and then the tape passes around the flywheel.

The tape passes on to the audio play head, which again has a narrow gap, and finally leaves the tight loop by passing the synchronous record head which gives the shingled recording previously described from low-voltage 60-cycle power. This new synchronous recording will match the spacing of the original recording because the latter is being used to hold the tape speed during this operation.

Speed Control

To control the tape speed, it is possible to use a synchronous motor driving the capstan, and change the frequency of the power driving the synchronous motor. But this takes considerable power, which is hard to generate from vacuum tubes, and a far more practical answer has been set up on this equipment, as shown diagramatically in Fig. 4. The synchronous motor is connected directly to the normal 60-cycle power. But the capstan is driven from the synchronous motor by means of a conical spindle on the motor, and a conical rubber drive wheel on the capstan shaft. So the actual rotational speed of the capstan is determined by the effective diameters of the motor conical spindle and the rubber drive, where the two contact.

Now, if the whole synchronous motor is moved up and down it is seen that the effective diameter of the conical spindle where it contacts the rubber drive will vary, and as a result the speed ratio between the capstan and the synchronous motor will change. A three percent plus or minus speed change from the 15-inch normal tape speed may be accomplished with the equipment as set up. The actual movement of the synchronous motor up and down is accomplished through a clockwork gear reduction from a very small a-c induction control motor. Or it may be accomplished manually, as shown by the speed control dial on Fig. 3 in the left center. A spring is used to counterbalance the weight of the synchronous motor.

A large operating switch is in the lower left hand corner of the top plate, Fig. 3, with positions as shown. In the position shown the control knob is at stop. Twisting it to the right puts the machine in play. In this position, the tape will play back a previously recorded signal and the speed of playback will be determined by the control signal. While the machine is in the play position the edit on-or-off switch may be flipped into the slot in the cam attached to the control knob, which then brings into action the edit switch on the lower right of the top plate.

This edit switch has three positions, center for stop, right for forward, ane reverse for backward. With this switch, the tape moves at normal speed forward and back which facilitates picking out spots on the tape, because their pitch will be normal even though the tape may be going backwards. A separate push button for recording (not shown) is then necessarily pushed to do recording, and a red light above the erase head shows that it is in operation, and a green light above the synchronous record head shows that it is putting the control track down on the tape.

For rewind, to take the tape back from the takeup reel on the right to the supply reel on the left, the control knob is turned to the position at the left of stop and then the knob on the far right lower part of the top plate may be used to change the speed of the rewind, so that selections, may be approached slowly when necessary. Then on turning the control knob to the extreme left the tape will go forward under this same right hand control knob, either slow or fast. The heads are assembled on swivel mounts, so that on rewind the mounts are turned back out of contact with the tape so that the tape does not wear them out unduly. The wear upon heads increases rapidly with tape speed. The rubber idlers in the center of the top plate are solenoid driven and are in position only when the control knob goes to play. This again relieves the tape tension on the idlers, capstan and heads on rewind.

Synchronous Playback

Finally there is the matter of picking the control signal off a recorded tape and making it hold the tape absolutely in step with the film, even though the tape may be stretched because of humidity or temperature changes. This is accomplished by changing the capstan speed, as previously described. The equipment that does it automatically is a synchronizer of special design, shown in Fig. 5. Starting with the synchronous playback head *a* in the upper left, the signal which should be 60 cycles from the

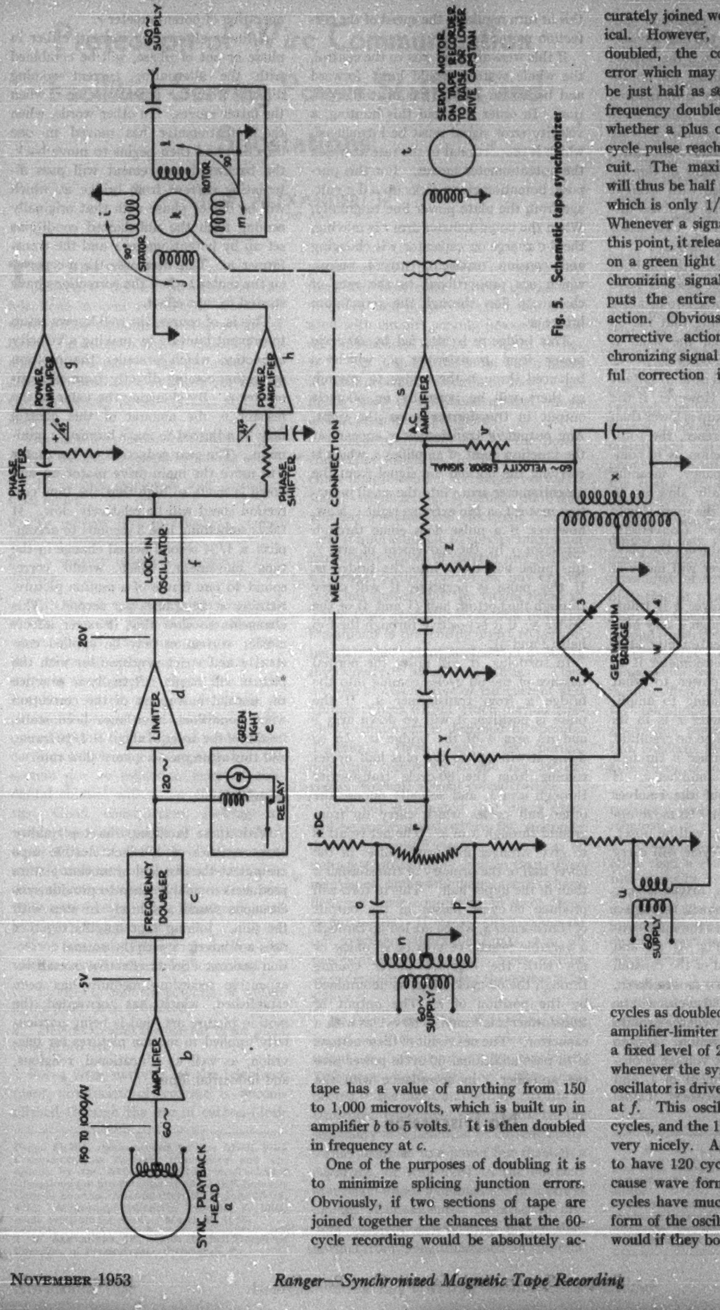

60~ SUPPLY
ROTOR
STATOR
90°
90°
k
l
i
j
m
POWER AMPLIFIER
g
POWER AMPLIFIER
h
+45°
-45°
PHASE SHIFTER
PHASE SHIFTER
LOCK-IN OSCILLATOR
f
20V
LIMITER
d
GREEN LIGHT
e
RELAY
120~
FREQUENCY DOUBLER
c
5V
AMPLIFIER
b
60~
150 TO 1000μV
SYNC. PLAYBACK HEAD
a
MECHANICAL CONNECTION
SERVO MOTOR ON TAPE RECORDER TO RAISE OR LOWER DRIVE CAPSTAN
t
A.C. AMPLIFIER
s
v
60~ VELOCITY ERROR SIGNAL
x
z
Y
GERMANIUM BRIDGE
1
2
3
4
w
+DC
r
o
p
n
60~ SUPPLY
u
60~ SUPPLY

Fig. 5. Schematic tape synchronizer

tape has a value of anything from 150 to 1,000 microvolts, which is built up in amplifier *b* to 5 volts. It is then doubled in frequency at *c*.

One of the purposes of doubling it is to minimize splicing junction errors. Obviously, if two sections of tape are joined together the chances that the 60-cycle recording would be absolutely accurately joined would be most problematical. However, if the frequency is doubled, the consequent out-of-phase error which may come from splicing will be just half as serious, because with the frequency doubled the effect is the same whether a plus or a minus half of a 60 cycle pulse reaches this point of the circuit. The maximum error in splicing will thus be half a cycle or 1/120 second, which is only 1/5 motion-picture frame. Whenever a signal from the tape reaches this point, it releases a relay *e*, which turns on a green light indicating that the synchronizing signal is present. This also puts the entire control operation into action. Obviously, there should be corrective action only when the synchronizing signal is present and a purposeful correction is indicated. The 120 cycles as doubled at *c*, then goes on to an amplifier-limiter at *d*, which establishes a fixed level of 20 volts up to this point, whenever the sync signal is present. An oscillator is driven by this 120-cycle signal at *f*. This oscillator is nominally at 60 cycles, and the 120 cycles input locks it in very nicely. Again it is advantageous to have 120 cycles for this junction because wave form variations in this 120 cycles have much less effect in the wave form of the oscillator at 60 cycles than it would if they both were at 60 cycles.

After oscillator *f*, the circuit is split with a phase shifter; the top half advancing the signal 45 degrees, and the bottom retarding it 45 degrees, using the appropriate capacitor-resistance combination. Power amplifiers *g* and *h* follow these phase shifts, and these, in turn, activate the 90-degree separated coils of the stator of the resolver *k*. The rotor of this resolver is, in turn, supplied with straight 60-cycle power, again with a phase differentiation accomplished by a capacitor in coil *l*, while coil *m* is fed directly.

These phase-shifting networks on both stator and rotor make the resolver have two rotating fields. If these two rotating fields are identical, meaning that the 60 cycles as picked up from the head is exactly in step with the 60-cycle power then being used, there will be no resultant motion of the resolver. However, if the frequency of the tape pickup is lower than that of the 60-cycle reference, the rotor will move in such a direction as to compensate for this discrepancy, meaning that the rotor will actually slow up to place it coils in line with the more slowly energized coils of the stator. Conversely, if the synchronizing signal from the playback is ahead, the resolver will move in the other direction.

The motion of this resolver *k* is transmitted mechanically, through a very high gear reduction, to a potentiometer at *r*. The purpose of this potentiometer is to apply normal 60-cycle power to what amounts to a bridge leading to amplifier *s*. When potentiometer *r* is in its center position, there will be no resultant voltage from the transformer *n* through the high pass filter *z* to amplifier *s*. If the mechanical motion of the resolver calls for the potentiometer *r* to move upward, the 60-cycle balance will be upset, and one phase of the 60 cycle will carry on through to amplifier *s*.

This amplifier, in turn, drives a very small induction motor *t* which, through a very high gear ratio, moves the main synchronous drive motor (Fig. 4) up and down to control the speed of the capstan on the top plate. If arm *r* moves down, the opposite phase of the 60 cycles will be applied to motor *t* and this will then rotate the motor in the other direction. It is to be noted that the 60-cycle voltage is proportional to the displacement of *r*, and this in turn regulates the speed of the correction motor *t*.

If this were all there was to the control, the whole system would hunt forward and backward about the neutral position. In order to avoid this hunting, a velocity-error signal must be introduced, which is proportional to the rate at which the potentiometer moves. For this purpose, potentiometer *r* picks up a d-c voltage from the plate power line to ground. When the potentiometer arm *r* is moving, the d-c charge on capacitor *y* is changing and causing transient current pulses, which are proportional to the rate of change to flow through the germanium bridge *w*.

This bridge *w* is also fed by 60-cycle power from transformer *u*, which is balanced through the bridge to ground, so there will be normally no 60-cycle output in transformer *x* on the right. Any output of transformer *x* appears at the junction point of amplifier *s*, which is carrying the normal a-c signal from the potentiometer arm *r* into the small induction motor *t* on the extreme right. Now, however, if a pulse does come through capacitor *y* by the movement of arm *r*, this pulse will carry into the bridge *w*. If this pulse is negative, it will carry through the bottom half (1 and 4) of the bridge *w;* if it is positive through the top half (2 and 3).

In so doing, it will upset the normal balance of the 60 cycles coming into the bridge *w* from transformer *u*. If the pulse is positive, it will go down arm 1 and up arm 4 of the bridge *w*. In so doing, it will add to the plus half cycles coming from the 60-cycle transformer through arm 1, and actually oppose the other half cycles which carry up from ground through arm 3. The net result is to give stronger half-cycle pulses in the lower half of the primary of transformer *x* than in the upper half. This in turn will produce 60-cycle pulses in the output of transformer *x*, which are led up through a variable resistance *v* to be in phase or not with the 60-cycle power coming through the 60-cycle filter, as determined by the position of *r*. The output of transformer *x* is tuned to 60 cycles with a capacitor. The net result of these actions is to pass additional 60-cycle power into the amplifier *s* in accordance with the operation of potentiometer *r*.

A linear alternating current, either in phase or out of phase, will be combined with the alternating current coming directly from the potentiometer *r*, when the latter moves. In other words, when the potentiometer has moved in one direction and then begins to move back, the backwards movement will pass alternating current from bridge *w*, which will be out of phase with that originally coming from the unbalanced conditions set up by potentiometer *r* and the transformer *n*. This decreases the a-c power on the motor *t* when the corrections have started to take effect.

This is, of course, the well known setup to prevent hunting by making a velocity correction which precedes the position correction coming directly from potentiometer *r*. By changing the value of the resistor *v*, the amount of this control may be adjusted to make hunting a minimum. The gear reduction on the motor *t* to move the main drive motor up and down is made so high that the tape correction speed will be relatively slow. It takes something like 3 seconds to accomplish a 1/24 second actual change in the tape movement, which would correspond to one frame of a motion picture, running at 24 frames per second. This change is so slow that it never affects music; correction may be applied constantly and strict synchronism with the picture will result. Actually in practice the normal movement of the correction when conditions have once been stabilized is of the order of about ±1/10 frame, and this again means a very slow rate.

Conclusion

With these facilities, the top quality characteristics of 1/4-inch flexible tape are put at the disposal of motion picture producers to enable them to provide synchronous sound absolutely in step with the film. Joining tape lengths together does not interfere with the normal correction process. So an effective overall inexpensive recording medium has been established, which has forwarded the motion picture art, and is being particularly applied to motion pictures for television, as well as educational, religious, and industrial films.

No Discussion

Protection of Wire Communication Facilities Serving Power Stations and Substations

T. W. ALEXANDER, Jr.
FELLOW AIEE

THIS DISCUSSION has to do with protective devices provided in the telephone system to furnish protection to telephone equipment and to personnel having occasion to use or maintain this plant. Particularly, the discussion will deal with protection available for use on circuits serving power stations and substations.

The primary reasons for installing protective devices in telephone plants are:

1. To protect employees and customers from injury;
2. To prevent fires in telephone central offices or customer's premises;
3. To protect telephone apparatus from damage.

With the exception of power used with certain carrier and radio equipment, where special procedures are in effect, the normal operating voltages and currents used in the operation of the telephone system are not dangerous to persons or property, but protection is required against lightning and potentials or currents which may be impressed on the system due to induction from or accidental contact with other electric facilities. Good construction, looking towards the maintenance of proper clearances from the facilities of other wire-using companies, is the first essential in our protective scheme. The use of protective devices provides a final safeguard to persons and property.

Two general types of protective devices are employed in telephone plant:

1. Devices to limit the potential or voltage in the plant to be protected;
2. Devices to limit the current in the plant to be protected.

For a large majority of the telephone plant, limitation of voltage is accomplished through the use of carbon-block protectors, and limitation of current by fuses and heat coils. Other special devices are used in specific cases, and those used with circuits serving power stations and substations will be discussed in detail. Fig. 1 shows the units which make up the protector assemblies most commonly used in telephone central offices and at customers' premises.

Typical systems in use by the power companies which use leased facilities include remote-control and supervisory systems, telemetering systems, pilot-wire relaying systems, and electronic devices.

It is impracticable to specify maximum permissible values of current and voltage for leased-facility operation, but experience has shown that the values listed in Table I are generally acceptable from the standpoint of preventing injury to personnel, damage to plant, and also of avoiding interference to telephone facilities. They are also representative of the current and voltage ranges within which normal telephone equipment is designed to operate. Therefore, these values for normal operating conditions should not be exceeded without special consideration.

It is not anticipated that many cases will be encountered where the customer desires to operate circuits at currents or voltages substantially higher than those mentioned, and many manufacturers have designed their equipment to meet these requirements. However, these values do not represent inflexible limits, and all pertinent facts in specific cases will be considered in determining the course to be followed.

In the many types of systems using leased facilities, a wide variety of terminal equipment may be employed. Telephone conductors are often connected to the input and output circuits of electronic equipment, connected to the secondary of a potential or current transformer, coupled to a power line through a capacitor, or connected to the contacts of a relay whose windings are in a power circuit. Consequently, in order to prevent voltages substantially higher than normal from being impressed on the telephone facilities, in the event of some equipment failure, it is important that any transformers, relays, or other apparatus connected between the telephone facilities and sources of power meet the requirements of the American Standards Association (dielectric strength of twice the line voltage, plus 1,000 volts). Apparatus listed by the Underwriters Laboratories meets these requirements.

In cases where the circuit arrangement or construction of the customer's equipment indicates that the apparatus may not meet the American Standards Association requirements, and that abnormal voltages are likely to be impressed on the telephone circuit, an isolating transformer having the necessary dielectric strength may be used. Carbon-block protectors having a separation of 3 mils between the line and ground blocks, and having a breakdown voltage of approximately 350 volts, installed between telephone conductors and ground at or near the customers' equipment may also be used. When leased conductors are classified as exposed, these protectors are required in any case. Telephone circuits are considered exposed when there is a possibility of contact with other circuits operating at more than 300 volts. As a rule, all telephone plant not in underground construction is considered exposed.

In general, the same protective arrangements would be applied to leased facilities and equipment as would be used for telephone circuits and equipment under the same conditions. An exception concerns the use of heat coils or 60-type fuses, as follows:

1. On leased circuits which merely loop through a central office, heat coils are not required;
2. Where a leased circuit enters a building other than a central office, the protection will normally be the same as for trunks serving a private branch exchange. However, the 60-type fuse may be omitted in cases where the terminal equipment directly associated with the leased circuits has no path to ground.

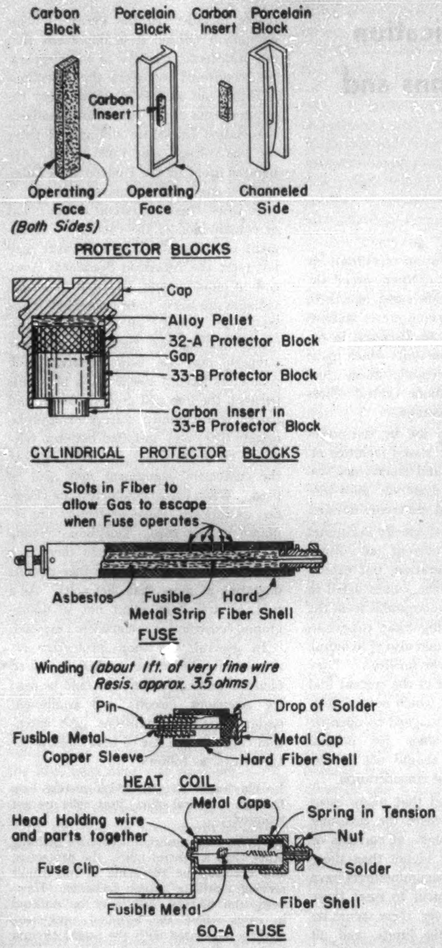
Stations at remote locations served by open wire such as at isolated substation locations, may be subject to service interruptions because of permanent operation

Table I. Maximum Permissible Voltages and Currents

	D-C	A-C (RMS)
Maximum voltage, conductor to conductor	270 volts with midpoint grounded	120 volts
Maximum voltage, conductor to ground	185 volts	120 volts
Maximum current in any conductor	0.35 ampere	0.35 ampere

Paper 53-203, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted February 4, 1953; made available for printing March 30, 1953.

T. W. ALEXANDER, JR., is with the Bell Telephone Company of Pennsylvania, Pittsburgh, Pa.


Carbon Block
Porcelain Block
Carbon Insert
Porcelain Block
Carbon Insert
Operating Face (Both Sides)
Operating Face
Channeled Side
PROTECTOR BLOCKS
Cap
Alloy Pellet
32-A Protector Block
Gap
33-B Protector Block
Carbon Insert in 33-B Protector Block
CYLINDRICAL PROTECTOR BLOCKS
Slots in Fiber to allow Gas to escape when Fuse operates
Asbestos
Fusible Metal Strip
Hard Fiber Shell
FUSE
Winding (about 1ft. of very fine wire Resis. approx. 3.5 ohms)
Pin
Drop of Solder
Fusible Metal
Metal Cap
Copper Sleeve
Hard Fiber Shell
HEAT COIL
Metal Caps
Head Holding wire and parts together
Spring in Tension
Nut
Fuse Clip
Solder
Fiber Shell
Fusible Metal
60-A FUSE

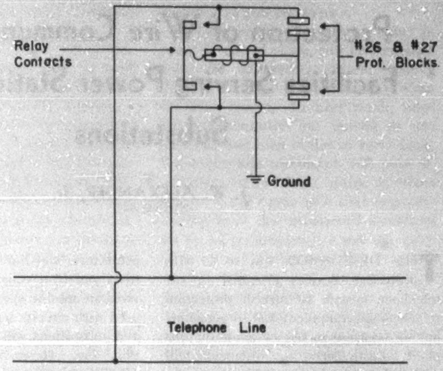

Relay Contacts
#26 & #27 Prot. Blocks
Ground
Telephone Line

Fig. 1 (left). Standard telephone protection equipment

Fig. 2 (above). Short-circuiting relay

of station protectors due to lightning surges. In these cases, consideration can be given to the use of additional protector blocks between the open-wire conductors and ground, at a point about $^1/_4$ mile from the station. These additional protector blocks would have a separation of 10 mils between the line and ground blocks, and would drain off the higher voltage surges and prevent them from being impressed on the station protector. The breakdown voltage of the 10-mil blocks is approximately 1,000 volts.

In some cases, the customer may wish to use his own station protective devices to protect his equipment, or he may even wish the telephone company to omit the station protectors entirely, on the basis that his equipment can withstand any likely abnormal potentials without failure or that it will interfere with the correct functioning of the circuit. Where all the equipment connected to the leased circuit is owned and maintained by the customer, arrangements such as those mentioned will be considered if adequate protection is assured for telephone personnel and plant in event of customer-equipment failure. However, in such cases, it will be important that a record of any special arrangements agreed upon be maintained. Where the telephone company owns and maintains any or all of the equipment involved, the protective equipment required will be determined after special consideration of all factors.

Grounding practices for protective equipment associated with leased facilities are the same as those normally applied to station protection. The order of preference is:

1. Public water system.
2. Public gas system.
3. Private water system.
4. Grounded metallic structures.
5. Multigrounded neutral power system electrode.
6. Ground rod.

It is, of course, desirable that the telephone protector employ the same grounding medium as that employed by the electric light or power service for the building. In cases where separate ground electrodes are used for the telephone and power services, it is desirable that they be bonded together so as to limit any difference in potential that may exist between grounded or protected parts of the two systems.

Certain circuits, because of the nature of the intelligence they transmit, require special protection considerations in addition to the leased-facilities protection as outlined above. An example of this type of circuit is the pilot-wire relaying circuit employed by power companies to sectionalize line faults. In this case, even momentary interruptions, caused by operation of protector blocks, may result in false operation of the power-line relaying and switching equipment, with corresponding serious interruption of the power circuits. Special protective arrangements, to be applied in any particular case,

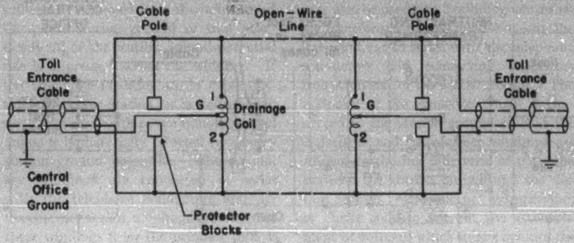

Cable Pole
Open-Wire Line
Cable Pole
Toll Entrance Cable
G
1
Drainage Coil
2
1
G
2
Toll Entrance Cable
Central Office Ground
Protector Blocks

Fig. 3. Direct drainage coils

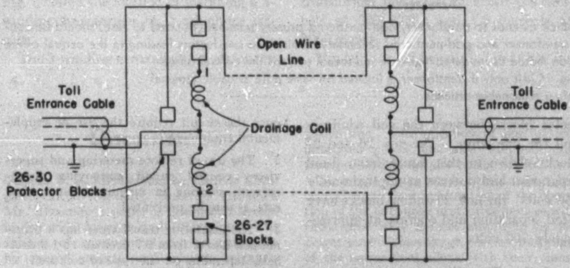

Open Wire Line
Toll Entrance Cable
1
Drainage Coil
2
Toll Entrance Cable
26-30 Protector Blocks
26-27 Blocks

Fig. 4. Direct drainage with additional protector blocks

depend upon the type of system to be operated over the leased facilities and whether such factors as ground potentials, low frequency induction, lightning, or a combination of these effects are significant. Some of the protective measures applicable in such situations include the use of:

1. Short-circuiting relay protector, to prevent permanent grounding of protector blocks;
2. Direct drainage, which practically eliminates even momentary operation of protectors;
3. Insulating transformer;
4. Neutralizing transformers.

Short-Circuiting Relay

A short-circuiting relay is a device used to place a short circuit around the blocks in an ordinary carbon-block protector within a very short time after the original discharge, thereby relieving the blocks of the duty of carrying the discharge current during a large part of the discharge interval. They are used principally on open-wire lines exposed to low-frequency induction.

Experience has shown that severe cases of induction may permanently ground protector blocks on exposed telephone circuits and cause interruption to service. Where the protector blocks are located along the telephone line or at outlying offices, it may take an appreciable amount of time for maintenance people to visit these locations and replace the carbon blocks. In these cases the interruptions to service may extend over a period of time which would be objectionable. The relay protector preserves the protective features of the carbon-block protectors and, at the same time, reduces very greatly the chance of permanent grounding.

The relay protector also provides substantial benefit from the standpoint of acoustic shock since, if the discharge current through the protector blocks is sufficient to operate the relay, both sides of the circuit are brought to the same potential after a very short time, and no circulating current results in the telephone circuit after the relay operation.

The relays in the short-circuiting protectors will short circuit the protector blocks in about 0.025 seconds after the initial block operates. Fig. 2 shows a schematic of the short-circuit relay protector, connected across carbon protector blocks on a telephone circuit.

Direct Drainage

During thunder storms, substantial voltages to ground may be induced on the conductors of an open wire line. The usual protectors provided at the ends of the line serve to relieve these high voltages by discharging them to ground. There is always, however, a certain amount of dissymmetry between the two protectors on a pair of wires, either in initial breakdown voltages, discharge characteristics, or both. Because of these differences, when the lightning currents discharge through the protectors, a pulse appears in the metallic circuit, whose magnitude depends upon the size of the voltage induced by the lightning and the degree of dissymmetry between the protectors.

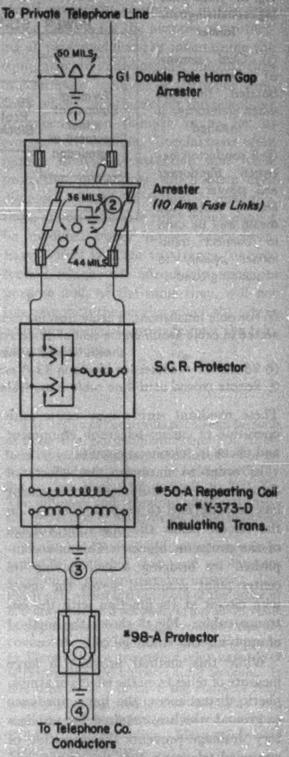

To Private Telephone Line
50 MILS
G1 Double Pole Horn Gap Arrester
①
Arrester
(10 Amp. Fuse Links)
36 MILS
②
44 MILS
S.C.R. Protector
*50-A Repeating Coil or *Y-373-D Insulating Trans.
③
*98-A Protector
④
To Telephone Co. Conductors

Fig. 5. Insulating transformer assembly

Separate leads shall be installed for grounds ① and ②. Grounds ③ and ④ may be connected to the same lead and all leads should be connected to the station ground
Where ground rods are used for grounds ① and ② these rods should be at least 15 feet from any other ground rod and at least 10 feet from each other

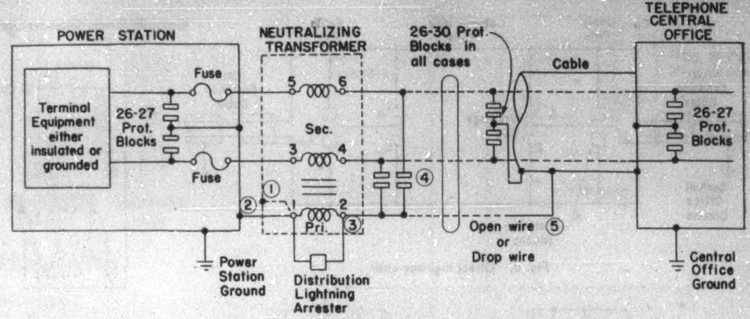


Fig. 6. Three-winding neutralizing transformer

① Odd terminal of primary to be bonded to case when transformer is installed

② If conductors between transformer and power station are in cable, the sheath may be used to connect transformer primary to station ground

③ For pole installations, if cable from the central office extends to transformer, odd-numbered primary terminal (bonded to case) should be connected to cable sheath on the central office side of transformer and odd-numbered secondary terminals to conductors leading to the central office sheath toward power station should be insulated from case and sheath toward the central office

④ 26 and 31-A protective blocks in 83-A mounting. Omit only if transformer is located on same pole as cable terminal

⑤ Remote ground used if no cable is available within a reasonable distance

These resultant surges may cause false operation of carrier-telegraph equipment and clicks in telephone circuits.

In order to minimize the effects of these surges, an equipment arrangement has been devised to keep the voltage on the circuit below the operating voltage of the protector blocks. This is accomplished by bridging a coil having its center point grounded across the open-wire circuit at its junction with the entrance cable. Fig. 3 shows the method of applying this drainage coil.

While this method provides a large measure of relief from the effects of atmospheric disturbances, the low impedance to ground which is required for satisfactory drainage prevents the operation of grounded telegraph over the circuit, and the use of d-c testing. It also affects the characteristic of voice-frequency repeatered circuits at the lower frequencies and introduces some modulation effects in the carrier system working over the particular pair. To obviate these disadvantages, an arrangement has been developed in which protector blocks are connected between the drainage-coil windings and the line wires. The drainage coil is thus kept off the circuit until a surge has broken down the protector gaps. Tests have shown this protector-drainage arrangement to be almost as effective as the direct-drainage arrangement.

Fig. 4 shows the arrangement of protector blocks between the coil windings and the line wires. The nos. 26 and 27 blocks shown in this figure have 3-mil separation and operate at approximately 350 volts; the nos. 26 and 30 blocks have 6-mil separation and operate at approximately 700 volts.

Insulating Transformers

Whenever an abnormal condition on a power system results in a current through the ground between a power station and some other point, the potential of the ground at the power station will differ from that of the distant ground and, if the difference is great enough, protection may operate on communication facilities serving the power station.

In order to guard against circuit outage due to protector operation at locations where it is expected that excessive voltages may be experienced, the communication circuits may be equipped with insulating transformers. Under this arrangement, a transformer is inserted in each telephone circuit at such location as to isolate the effect of the power-station ground from any distant ground. Gap protectors and fuses are installed on the line side of the transformer and standard station protection on the station side.

The disadvantages of the insulating transformer arrangement are:

1. Direct-current signaling and supervision over the circuit require the use of supplemental equipment;
2. The use of remote metering and supervisory control circuit employing direct-current, varying in magnitude at a slow rate, is usually prevented;
3. The insulating transformer has a transmission loss of from 0.7 decibels to 1.2 decibels, depending on the type of coil used;
4. Reduces ringing range.

The insulating transformer may also be used to protect telephone facilities against voltages which may be impressed on them by highly exposed circuits of the customer, as for example circuits routed on higher voltage transmission lines. In these cases, transformers are usually installed at the junction of the customer owned and telephone company circuits. Usual types of transformers have breakdown voltages of approximately 25,000 volts between primary and secondary windings, and between primary or high winding and the case. Fig. 5 shows the arrangement of this transformer.

Neutralizing Transformers

Neutralizing transformers are also used to protect against excessive ground potential at a power station. A 3-winding neutralizing transformer is one having one primary winding and two secondary windings. The primary winding should be connected between the power-station ground and an outside ground, which is

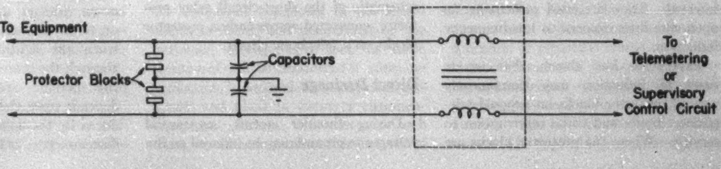


Fig. 7. Two-winding neutralizing transformer

sufficiently removed from the influence of the power-station ground so that practically all of the station ground-potential rise is impressed across the winding. If the telephone facilities are in cable outside the power station, it is desirable to use the cable sheath as the remote ground unless it is materially affected by power-station ground potentials. The secondary windings are connected in series with the telephone conductors and, by transformer action, a voltage is induced in these windings from the primary that is nearly equal and opposed to the voltage directly impressed on the conductors. Fig. 6 shows the method of installing a 3-winding transformer on a telephone circuit.

The 3-winding neutralizing transformer may also be used as a 2-winding neutralizing transformer by making use of only the line windings in series with the telephone conductors and connecting capacitors from each side of the line to ground. The midpoint of the capacitor circuit is connected to the station ground so that the capacitors and transformer windings are subject to the full difference of potential between the station ground and the distant ground. The coils act as impedance or choke coils to prevent the establishment of voltages which might operate the carbon protectors. These connections are shown in Fig. 7. The neutralizing transformer has a loss of about 0.8 decibels at 200 cycles and 1.6 decibels at 3000 cycles.

It again should be pointed out that the use of protective equipment in the telephone plant is for the purpose of protecting all personnel using or maintaining the equipment, to guard against starting fires in customers' premises or telephone company offices, and to protect the telephone equipment from damage in so far as economically feasible. Special cases often require individual attention and while protective procedures which have been discussed should be followed generally, there are no inflexible rules which cannot be modified under proper supervision. Every effort is made to arrange for protection which will adequately serve its purpose and, at the same time, will not interfere with the service which the power company wishes to provide over the telephone facilities.

◆

Discussion

Gordon Y. R. Allen (Osborne Electric Company, Ltd., Toronto, Ontario, Canada): Mr. Alexander's paper is one which should invite wide interest and discussion. The subject of communications protection in its many aspects and phases certainly warrants more technical papers. In general, I agree with the author's analysis, although I do feel that these problems need some elaboration.

Basically, the paper appears to be written from the viewpoint of leased telephone circuits, whereas the problems are equally common to electric-power utility-owned and -operated wire circuits. It is apparent that the use of the protective devices detailed and illustrated in the paper is Bell Telephone practice, and no mention is made of the several other common protective devices widely used. Particularly I am concerned with the omission of the Osborne telephone protector which is in common use in the United States, Canada, and the Latin American countries, as well as in many other foreign countries.

In the list of protective devices in use today, the combined isolating and drainage transformer is not indicated. This type of transformer has been in common use, stretching over a period exceeding 20 years. It forms the heart of the Osborne telephone protector. This transformer, along with the associated apparatus, is shown in Fig. 8, and it is interesting to recall that patents were issued to R. W. Osborne many years ago on this device.

Referring for the moment to the question of induction, it is important to remember that there are two types, that is, the electromagnetic and electrostatic or, in other words, voltage induced between the conductors of the telephone line, and voltage induced between the conductors and ground. The voltages and currents are not necessarily of the same magnitude, and it is often necessary to analyze these components. This is particularly the case where a telephone company jointly uses the same poles as a power line for varying distances. One therefore must consider the characteristics of the power line. The question of possible co-ordinated power line and telephone line transpositions must also be considered.

In recent times much consideration has been given to the co-ordination of the operation of the telephone protective devices with those protective devices employed on the power lines. The voltage, current, and phase of the power lines, along with their circuit breakers or fuses, determine to a great extent the time voltage rise and resulting currents on the telephone lines. The telephone protective devices therefore must be able to withstand the fault voltages and currents prior to the time at which the power line protective devices operate or the fault is cleared. At least the telephone protective devices, if expendable, must prevent voltage from appearing across terminal apparatus.

With reference to the use of an isolating transformer, I call attention to an important difficulty in the case of a fault. If a telephone line becomes unbalanced or grounded on one side, the transverse impedance of the line remains unchanged. Any extraneous potential in the grounded side, for instance, a power line conductor becoming grounded, is impressed on the equipment in shunt with the line. If an insulating transformer is interposed between the line and telephone, it transforms the transverse potential to the saturation point of the transformer core, above which excessive exciting current will damage the primary winding. The voltage across the transformer secondary may be very high, depending on the frequency of the extraneous potential.

With respect to the use of a drainage coil alone, I agree with the limitations where direct current is involved. It is not difficult to protect the a-c apparatus involved but it is difficult to protect adequately the d-c equipment. In any event, under severe disturbances, unless both ends of the drainage coil are simultaneously grounded, it will simply burn out because of its inability to sustain heavy currents to ground. Carbon-block protectors will not protect a drainage coil where severe faults are encountered. In actual practice, I have seen a drainage coil which exploded when a 110-kv conductor came in contact with a telephone line at a distance of 1 mile.

The design of the drainage coil can be such as to obviate the complaint regarding its effects on voice frequency and carrier. For instance, bridging loss can be held practically flat at less than 0.02 decibels to 200 Kc. The drainage coil therefore offers only partial protection in many cases, while in others it may be adequate.

The short-circuit or grounding relay mentioned is somewhat slow, and no data are given on its current carrying capacity. Unless it can withstand currents of the order of 500 amperes for 3 seconds, it has little value in applications where high voltage lines are concerned. The Osborne Electric Company in Toronto has a grounding relay capable of withstanding 650 amperes for 3 seconds. It has been known to have the contacts completely disintegrated where the relay was used to protect a telephone line exposed to a 220-kv transmission line. More will be said about this relay later.

I should like to suggest an important personnel hazard which was not mentioned in the paper in connection with the lineman who uses a portable telephone on exposed open-wire circuits. Usually these portable telephones are fitted with carbon-block protectors which in many cases are most inadequate. Far better protection can be had with the use of an isolating transformer which could be considered expendable if damaged as a result of overexcitation. The Osborne company manufactures a 10-kv insulated plastic-moulded portable isolating transformer weighing only $2^{1}/_{2}$ pounds. The transformer has a very low bridging and transmission loss and is a ring-through type for use with a magneto.

To circumvent the difficulties of straight isolating transformers, drainage coils, and grounding relays used separately or independently, a combined isolating and drainage transformer, as mentioned previously, has been developed. This transformer is the heart of the Osborne telephone protector. When such a transformer is interposed between line and telephone, the neutral point of the primary or line winding is connected to ground. This lowers the impedance to ground to a very low value without normally altering the transverse impedance. Any unbalance in the telephone line, or

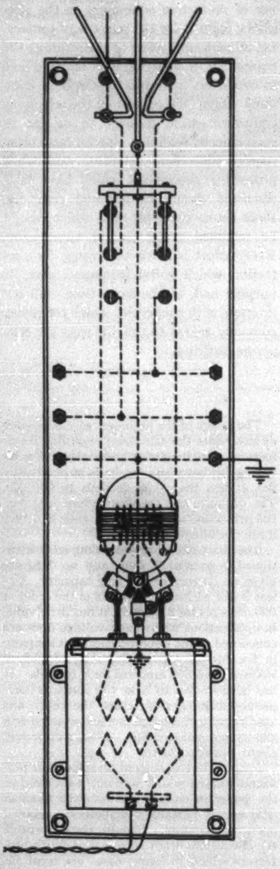


Fig. 8 (left). Type-DC telephone protector

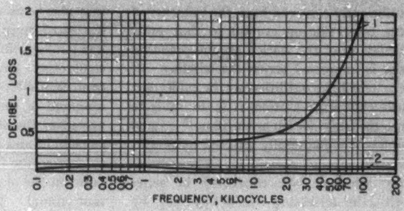


Fig. 9 (right). Osborne telephone transformer

Curve 1—Insertion loss of carrier transformer

Curve 2—Bridging loss of carrier transformer (at 6.66 megawatts on 600 ohms)

ground to one side, affects the total impedance to ground only negligibly, but greatly lowers the transverse impedance. This is because the transformer neutral ground and line ground constitute a short circuit of one half of the primary winding, and the transverse impedance therefore becomes the impedance of the transformer under such short-circuit conditions. The design of Osborne transformers is such that the short-circuit impedance is very low, so that the transverse voltage appearing across the secondary winding will be of low value. To maintain this condition, any circuit opening or short-circuiting device must operate on both sides of the circuit simultaneously. Any transformer or drainage coil has a certain capacity, and so it is with the drainage capacity of the Osborne transformer, which will drain 3 amperes continuously. However, to protect this transformer, a grounding relay is used which will short-circuit the line to ground when current through the transformer exceeds its capacity.

The Osborne grounding relay is of a repulsion-induction design for gaining fast operation, as little chatter as possible, and a quick release. Residual magnetism has no effect on the relay, and the relay has the same operating characteristic from power line fundamental to harmonics. The operating force is linear throughout the travel of the armature and is slightly greater when the armature is open. This relay will handle 650 amperes for a period of 3 seconds without damage, and the relay coils serve as chokes to facilitate discharge at a set of horn-gap arresters.

Fig. 8 shows the complete protector, while Fig. 9 shows the characteristics of the transformer. Note the very low transmission and bridging losses as compared to the losses of the 0.7 to 1.2 decibels mentioned in the paper.

Recently, the Osborne company designed a system of protection to serve a line carrying voice-frequency 4-channel carrier, and railway d-c block signalling. The line is in joint use with a 23-kv single-phase power line 360 miles long. This system will go into operation shortly. It is installed in the far North of Canada, where it is subjected to severe climatic conditions.

H. T. Seeley (General Electric Company, Philadelphia, Pa.): From the standpoint of the maker or the user of pilot-wire relays, the policies set forth in this paper represent a distinct advance over those frequently encountered in the past. The regional telephone companies have often required carbon-block protectors connected to the station ground from the line side, either of insulating transformers or neutralizing transformers, and therefore the pilot wires have become short-circuited at the only moment when the pilot-wire relays have a chance to justify their installed cost of $2,000 or more for a pair of terminals. The result is that the relays, according to one scheme, are useless because they fail to trip when they should; or that, according to the other scheme, they trip when they should not, which would render them worse than useless, in the opinion of most power companies. This pilot-wire trouble is one of the reasons why the operating record of pilot-wire relaying[1] is the poorest of any modern form of power-system relaying, and the record on pilot circuits that are leased is worse than that on such circuits that are privately owned.

The new policies are encouraging to the manufacturer because they open the possibility of taking advantage of a 1-second or 2-second rating for maximum current and voltage. They are encouraging to both the manufacturer and the user because they open the possibility of designing the pilot channel protection so that it will not operate for longitudinally induced potentials and ground-potential difference, but only for direct lightning strokes or crosses with power circuits. I should like to have the author advise to what extent these more liberal policies are than those of the American Telephone and Telegraph Company, so that we may know what may be expected in other areas than his own.

The ideal pilot circuit from the viewpoint of the relay engineer has been described recently[2] as one insulated to withstand 20 kv to ground for 1 minute, or 1.5 kv between conductors momentarily; this is being installed for several circuits, some as long as 16 miles.

While we can hardly hope to meet this high level simply by proper design and application of special terminal equipment on facilities designed primarily for other uses, the new policies give promise of meeting the requirements of the majority of pilot-wire relay installations.

REFERENCES

1. Pilot-Wire Circuits for Protective Relaying—Experience and Practice, 1942–1950, **AIEE Committee Report.** *AIEE Transactions*, vol. 72, pt. III, April 1953, pp. 331–36.

2. A Pilot-Wire Relaying Scheme Utilizing the Product Differential Relay, **R. I. Ward, D. W. Gilman.** *AIEE Transactions*, vol. 72, pt. III, Oct. 1953, pp. 911–17.

T. W. Alexander, Jr.: While our practices require the carbon-block protectors on the line side of the transformer, it should be pointed out that the ground associated with these blocks is remote ground and not station ground.

Also, it should be noted that the omission of carbon-block protection applies in cases where all the equipment connected to the leased circuit is owned and maintained by the customer, and adequate protection is assured for telephone personnel and plant in the event of customer equipment failure.

Where telephone company equipment is connected to the facility, carbon-block protection will in general be required, since that form of protection best meets our requirements. However, these blocks should not operate unless the facility is disturbed by lightning or contact with power circuits high enough in magnitude to operate the lightning arrestor protecting the neutralizing transformer.

The protection arrangement set forth in this paper was developed with the concurrence of the American Telephone and Telegraph Company and should be acceptable in all locations except where local conditions require modification.

The Map Method for Synthesis of Combinational Logic Circuits

M. KARNAUGH
NONMEMBER AIEE

THE SEARCH for simple abstract techniques to be applied to the design of switching systems is still, despite some recent advances, in its early stages. The problem in this area which has been attacked most energetically is that of the synthesis of efficient combinational that is, nonsequential, logic circuits.

While this problem is closely related to the classical one of simplifying logical truth functions, there are some significant differences. To each logical truth function, or Boolean algebraic expression, there corresponds a combinational circuit which may be constructed from a given set of appropriate components. However, minimization of the number of appearances of algebraic variables does not necessarily lead to the most economical circuit. Indeed, the criteria of economy and simplicity may vary widely for different types of components. A general approach to circuit synthesis must therefore be highly flexible. What is perhaps most to be desired is a simple and rapid technique for generating a variety of near-minimal algebraic forms for the designer's inspection.

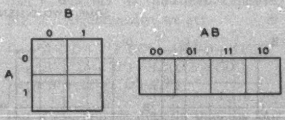
Boolean algebra,[1] or the calculus of propositions, is a basic tool for investigation of circuits constructed from 2-valued devices. Its direct application to synthesis problems is, nevertheless, not completely satisfactory. The designer employing Boolean algebra is in possession of a list of theorems which may be used in simplifying the expression before him; but he may not know which ones to try first, or to which terms to apply them. He is thus forced to consider a very large number of alternative procedures in all but the most trivial cases. It is clear that a method which provides more insight into the structure of each problem is to be preferred. Nevertheless, it will be convenient to describe other methods in terms of Boolean algebra. Whenever the term "algebra" is used in this paper, it will refer to Boolean algebra, where addition corresponds to the logical connective "or," while multiplication corresponds to "and."

The minimizing chart,[2] developed at the Harvard Computation Laboratory, represents a step in the desired direction. It makes possible the fairly rapid derivation of near-minimal 2-stage forms. By a 2-stage form is meant a sum of products of the elementary variables, or else a product of sums of the elementary variables. These expressions may then be further reduced by algebraic factoring. The chief drawback to this method lies in the necessity of writing, and perhaps erasing, on a chart that, for n variables, contains 2^{2n} entries. Thus, we must keep track of 1,024 entries for five variable problems and 4,096 entries for six variable problems.

E. W. Veitch[3] has suggested a method whereby results similar to those yielded by the minimizing chart can be obtained from an array containing only 2^n entries in a more rapid and elegant manner. The map method, which is explained in this paper, involves a reorganization of Veitch's charts, an extension to the use of 3-dimensional arrays, and some special techniques for diode and relay circuits.

Maps

Let the active and inactive conditions of the inputs to a combinational circuit be designated by assigning the values 1 and 0 respectively to the associated algebraic variables. An assignment of a simultaneous set of values to the n variables for a given problem will be called an input condition. There are 2^n possible input conditions.



(A) (B)

Fig. 1. Graphical representation of the input conditions for two variables

(A) Along two axes
(B) Along a single axis

For example, with only two variables, there are four input conditions. They may be represented graphically by the four squares in Fig. 1(A). Here, the values of variables A and B have simply been plotted along two perpendicular axes. It should be noted that squares which are adjacent, either horizontally or vertically, differ in the value of only one of the variables.

If Fig. 1(A) is cut along its horizontal midsection and the bottom half is rotated into line with the top, as in Fig. 1(B), then a representation of the input conditions for two variables is obtained along a single axis. Let us consider the squares at opposite ends of the row to be termed adjacent, as if it were inscribed on a cylinder. Then, as before, adjacent squares differ in the value of only one variable. Conversely, if two input conditions differ in the value assigned to just one of the variables, they are represented by adjacent squares.

If one also makes use of the vertical axis, one can represent the input conditions for three variables as in Fig. 2(A), and for four variables as in Fig. 2(B). In the latter case, opposite ends of each row or column should be considered adjacent, as though the figure were inscribed on a torus.

The labels on the diagrams may be simplified as shown in Fig. 3. The rows or columns within a bracket are those in which the designated variable has the value 1, while it is 0 elsewhere.

A combinational circuit of the type under consideration has a 2-valued output which is a function of the input condition. The synthesis problem may be said to begin with the specification of this functional dependence. Such information may be represented on a map as follows: Place a 1 in each square which represents an input condition for which the output is to have the value 1. The other squares may be imagined to contain zeroes.

Synthesis of 2-Stage Forms

Consider the function mapped in Fig. 4(A). Its algebraic realization is the product $A'BC'D$, where the primes indicate negation or complementation, for $A'BC'D=1$ if, and only if, $A=0$, $B=1$, $C=0$, and $D=1$.

Let us define a complete product to be a product in which each of the variables appears as one factor, either primed or

---

Paper 53-217, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 17, 1953; made available for printing April 23, 1953.

M. Karnaugh is with the Bell Telephone Laboratories, Inc., Murray Hill, N. J.

The author wishes to express his indebtedness for many valuable suggestions, help, and encouragement to E. F. Moore, K. Goldschmidt, and W. Keister, all of the Bell Telephone Laboratories.

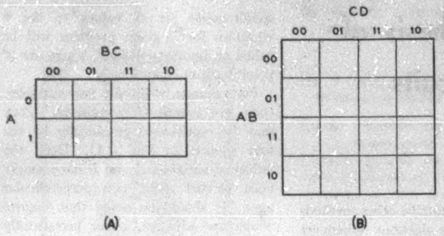


Fig. 2. Graphical representations of the input conditions for three and for four variables

not. Then any function whose map contains a single 1 may be represented by a single complete product. Each factor is primed if, and only if, it has the value 0 at the square in question. Because each square that contains a 1 gives rise to a product, such squares will be called *p*-squares.

If the map of a function contains *k* *p*-squares, then the function may be represented by the logical sum of the corresponding *k* complete products, each selected by this rule. This form of representation is the complete disjunctive normal form of the calculus of propositions. It is often the starting point for algebraic simplification.

However, it is usually possible to write down a more economical representation than a complete normal form by direct inspection of a map. Consider the function whose map is shown in Fig. 4(B). Its complete disjunctive normal form is $AB+A'B$. This is easily reducible algebraically: $AB+A'B=B(A+A')=B$. Now note that the *p*-squares on the map are precisely that set for which $B=1$.

Let us define a subcube to be the set of all squares on a map over which certain of the variables have fixed values. A subcube formed entirely of *p*-squares will be called a *p*-subcube.

Each *p*-subcube may be regarded as the map of a product formed according to the rules:

1. The factors of the product are those variables whose values are fixed within the subcube.
2. A factor is primed if, and only if, its value within the subcube is 0.

Fig. 5 shows some typical *p*-subcubes and the corresponding products. Each *p*-subcube may be thought of as a simply connected square or rectangular group of *p*-squares, if it is recalled that opposite ends of columns and rows are adjacent.

If *m* variables are not fixed in a given subcube, it is said to be *m*-dimensional, and it contains 2^m squares. A single square is thus a zero-dimensional subcube. Note that the larger *p*-subcubes correspond to products having fewer factors, since fewer variables are fixed in them.

It is now easy to see how to obtain economical 2-stage forms from maps. The rules are:

1. Choose a set of *p*-subcubes which includes every *p*-square at least once. In general, it is desirable to make the selected subcubes as large and as few in number as possible.
2. Write down the sum of the products which correspond to the selected *p*-subcubes. This gives the desired expression.

As an example of this procedure, we can, for the function mapped in Fig. 6, make the selection

$$f=AC'+A'CD+BCD$$

An alternate procedure is possible that leads to a product of sums, that is, a conjunctive normal form instead of a sum of products. First, this procedure is used to obtain an expression for the negative of the function mapped. This is done by considering the empty squares to be the new *p*-squares. In the case of Fig. 6

$$f'=A'C'+CD'+AB'C$$

The function desired, which is the negative of this, is now obtained by the simul-

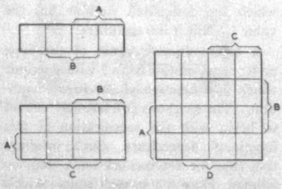


Fig. 3. Input representations with simplified labels

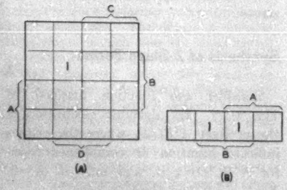


Fig. 4. Maps of two functions

(A) $f=A'BC'D$
(B) $f=A'B+AB=B(A'+A)=B$

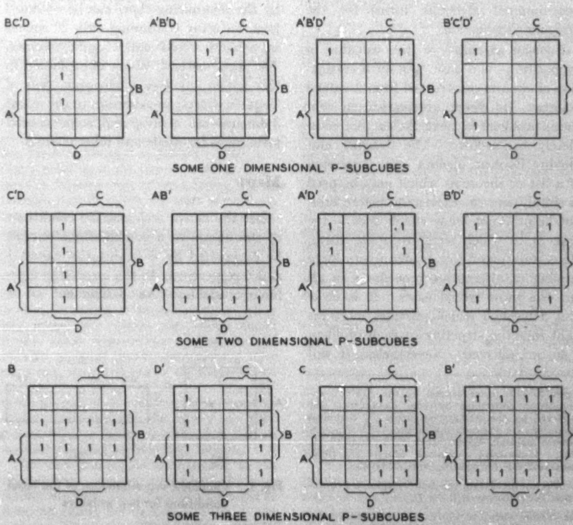

BC'D
A'B'D
A'B'D'
B'C'D'
C
B
A
D
SOME ONE DIMENSIONAL P-SUBCUBES
C'D
AB'
A'D'
B'D'
SOME TWO DIMENSIONAL P-SUBCUBES
B
D'
C
B'
SOME THREE DIMENSIONAL P-SUBCUBES

Fig. 5. A number of typical p-subcubes and the corresponding algebraic products

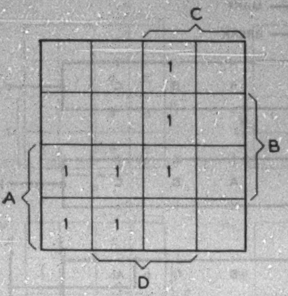


Fig. 6. Map of a function

taneous interchange of primes and nonprimes, and of multiplication and addition signs.

Thus

$$f=(A+C)(C'+D)(A'+B+C')$$

Both of these procedures have been proposed by Veitch.[3]

Minimal 2-Stage Forms

In combinational diode circuits, there is usually one diode per input lead to every stage. For 2-stage circuits, this means one diode per appearance of each algebraic variable plus one diode per product, or per sum, of these variables. It is often a simple matter to minimize rigorously the number of diodes used in such a circuit.

Consider Fig. 7(A). The dotted lines correspond to the choice of p-subcubes.

$$f=B+AC$$

Now note that asterisks are placed in two of the p-squares, so chosen that no single p-subcube includes both of them. Hence at least two p-subcubes are required. Furthermore, the selected p-subcube containing each asterisk is of maximum possible dimensionality. Hence each of the corresponding products contains the minimum number of factors.

The same kind of proof must be carried out for the alternate procedure, as illustrated in Fig. 7(B). Here we have

$$f'=A'B'+B'C'$$
$$f=(A+B)(B+C)$$

This is not as good, however, as the previous result, which we have now proved to be minimal in

1. Number of terms
2. Appearances of the variables
3. Diodes

This proof depends upon the fact that no one pair of asterisks lies in the same p-subcube. In some cases it may be found that only k asterisks can be placed on a map in this manner, and yet more than k terms are required to represent the function. When this occurs, a proof that at least $k+1$ terms are necessary can be carried through by contradiction. When the attempt to associate a p-subcube with each asterisk is made, it will be found impossible to include all p-squares in the k p-subcubes so selected.

Factoring by Inspection

When circuits are not restricted to the 2-stage variety, it is sometimes advantageous to reduce further the 2-stage forms by algebraic factoring. It is of some importance to show that factoring may also be carried out directly by inspection of a map.

For example, the function mapped in Fig. 8 is

$$f=A'B'+B'C=B'(A'+C)$$

Since both the chosen p-subcubes lies within subcube B, the presence of the common factor is established by inspection.

Occasionally, observation of the possibilities for factoring will determine the selection of subcubes and lead to a better circuit than would otherwise be obtained. In the case of Fig. 6, the choices

$$\begin{aligned} f&=AC'+A'CD+BCD=AC'+CD(A'+B)\\ f&=AC'+A'CD+ABD=A(C'+BD)+A'CD\\ &\quad\text{or}\qquad\qquad\qquad\;\;=AC'+D(A'C+AB)\end{aligned}$$

lead to equally good 2-stage forms; but the former yields the best factored form. Inspection of the map indicates that p-subcube BCD lies in CD along with $A'CD$, thus providing two common factors, while the alternative choice of ABD will give only a single common factor in either of two ways. When inspecting the map, it is not necessary to think of these subcubes by name as we must in the text, but merely to observe their relations, as sets of p-squares.

Even more extensive use of the set theoretic union (our $+$) and intersection (our $\cdot$) relations is possible. Consider Fig. 9. Algebraically, we get

$$\begin{aligned} f&=A'B'C'D+A'B'CD'+ABC'D+ABCD'\\ &=A'B'(C'D+CD')+AB(C'D+CD')\\ &=(A'B'+AB)(C'D+CD')\end{aligned}$$

But it can be seen directly that the four p-squares form the set which is the intersection of the union of $A'B'$ and AB and the union of $C'D$ and CD'. Thus $f=(A'B'+AB)(C'D+CD')$, as illustrated by the dotted lines.

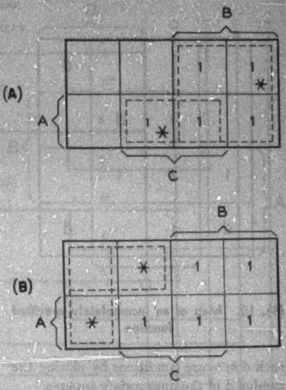


Fig. 7. Maps used to minimize a diode circuit

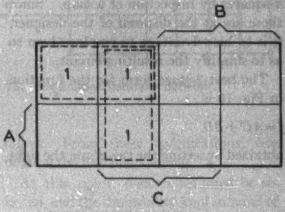


Fig. 8. Map of a factorable function

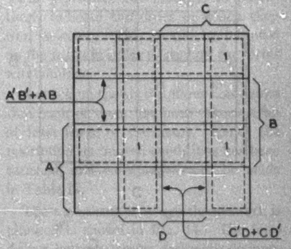


Fig. 9. Set theoretic interpretation of a map

"Don't-Care" Conditions

Very often, the output of a circuit is subject to less rigid restriction than the assignment of a definite value, 0 or 1, for some input conditions. The simplest such case is that of no restriction at all This may occur because the input conditions in question never are realized in practice, or because the output has no effect in those cases. We shall designate

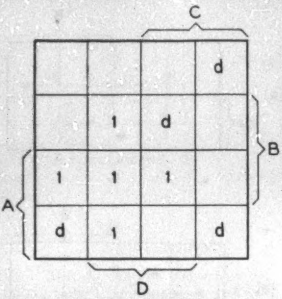


Fig. 10. Map of an incompletely specified function

such don't-care conditions by placing the symbol d in the appropriate squares.

It is usually quite simple to make an economical assignment of values to the d-squares by inspection of a map. Since these are at the disposal of the designer, it is to his advantage to employ them so as to simplify the resulting circuit.

The best 2-stage form for the function in Fig. 10 is

$$f = AC' + BD$$

obtained by setting the two d's on the right equal to 0, and the other two equal to 1.

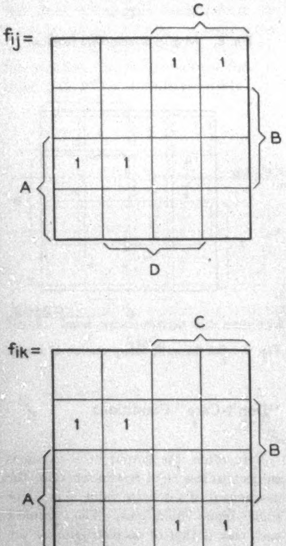


Fig. 11. A 2-output problem

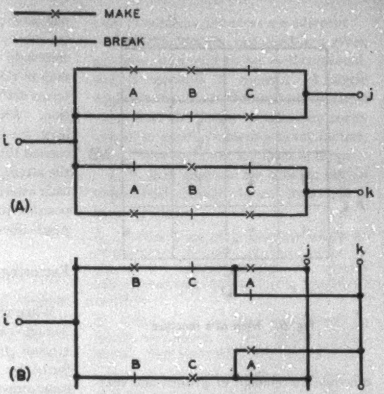


Fig. 12 (right). Synthesis of a 2-output circuit

(A) Two separate contact networks

(B) Networks disjunctively combined

The rule for making such choices is as follows: Assign values to the d's which enlarge and combine the necessary p-subcubes as much as possible but do not make necessary the selection of any additional subcubes.

The ease with which don't-cares can be properly evaluated is one of the major advantages shared by the minimizing chart, Veitch chart, and map methods in varying degree.

Disjunctive Combination in Relay Nets

The map method, inasmuch as it yields expressions in Boolean algebra, can be used to design 2-terminal, series-parallel relay contact networks, but not bridge-type 2-terminal networks. Hence, many 2-terminal contact networks designed by means of the map method will not be minimal in contacts or springs. This will be true, in particular, of the symmetric circuits.[4]

However, in the case of complicated, multioutput networks, the map method may be a very effective tool. Suppose that terminal i is a ground, to be connected through networks f_{ij} and f_{ik} to the output terminals j and k respectively. The specifications for f_{ij} and f_{ik}, which are networks on the contacts of relays A, B, C, D, are mapped in Fig. 11. If each net is synthesized separately, there results the circuit of Fig. 12(A). In Fig. 12(B), it is shown how, with a slight rearrangement, parts of the upper paths to j and k can be combined, as can parts of the lower paths. This results in a saving of four contacts.

The second circuit is completely equivalent to the first, for the transfers on relay A prevent any sneak paths between terminals j and k. While disjunctive combinations of this sort are certainly not new to the relay art,[5] this section is included to show how they may easily be recognized on maps, and hence how they play a part in the selection of subcubes.

Note that the paths ABC' and $A'BC'$, which give rise to one of the combinations, differ by only a prime on A. The corresponding subcubes in Fig. 11 are seen to be related by a simple displacement. The same is true for the other pair of p-subcubes.

A little practice will enable the designer to evaluate the various possibilities for factoring and disjunctive combination by inspection of the maps. It will then be a simple task to make a good choice of p-subcubes.

Unnecessary Contacts

It is of interest to note that for any given function some of the variables or their primes may be unnecessary. That is, it is possible to find an algebraic representation of the function in which these variables, or negated variables, do not appear. Hence the corresponding relay

Table I. Specifications for a Coded Decimal Digit Translator

Digit	1	2	4	5	Z	O	T	F	S
0	0	0	0	0	0	0	0	1	1
1	1	0	0	0	1	1	0	0	0
2	0	1	0	0	1	0	1	0	0
3	1	1	0	0	0	1	1	0	0
4	0	0	1	0	1	0	0	1	0
5	0	0	0	1	0	1	0	1	0
6	1	0	0	1	0	0	1	1	0
7	0	1	0	1	1	0	0	0	1
8	1	1	0	1	0	1	0	0	1
9	0	0	1	1	0	0	1	0	1

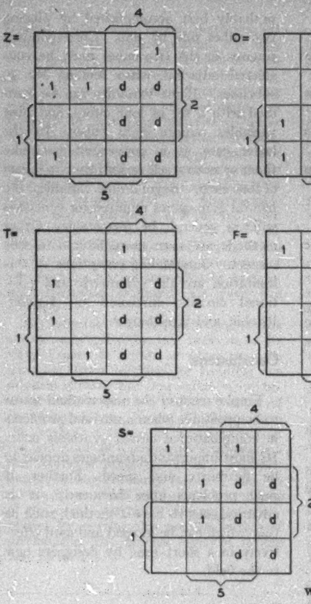


Fig. 13. A translator problem

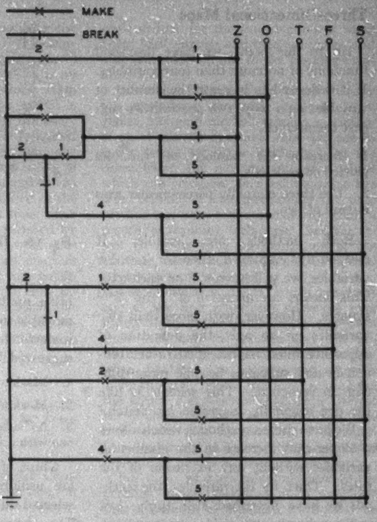


Fig. 15. The finished translator network

contact network will not contain make-contacts, or break-contacts, on some of the relays.

For example, the functions in Fig. 11 are shown on four-variable maps, but they may be realized in terms of only three variables, as in Fig. 12. Neither D nor D' is necessary.

In this case, it can be seen at a glance that the patterns appearing in the D and D' subcubes in both maps are identical. Therefore the output is independent of the value assigned to D. This is a case wherein both the following rules hold:

1. A function may be represented without the appearance of an unprimed variable, say D if, and only if, to each p-square in subcube D there corresponds an adjacent p-square in subcube D'.
2. A function may be represented without the appearance of D' if, and only if, to each p-square in subcube D' there corresponds an adjacent p-square in D.

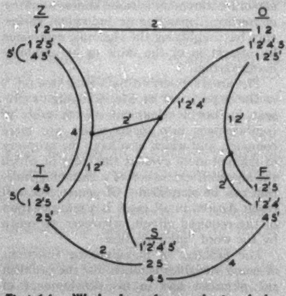


Fig. 14. Work sheet for synthesis of the translator

Illustrative Example: A Relay Translator

Suppose it is desired to find a relay contact network to translate coded decimal digits from a 1-2-4-5 code to 2-out-of-5 code. The five outputs will operate the relays Z (zero), O (one), T (two), F (four), and S (seven). The required translation properties are listed in Table I. The unarithmetic representation for zero is standard in the 2-out-of-5 code.

The remaining six input conditions for the 1-2-4-5 relays are unused or don't-care conditions. However, it is required that none of these conditions results in operation of zero or two of the five output relays. From these specifications, one obtains the five maps in Fig. 13.

At this point, p-subcubes must be selected, and the desirability kept in mind of factoring and disjunctive combinations. The chosen p-subcubes are listed in Table II, where the numbers in parenthesis indicate the order in which they were selected. This should be followed on the maps in order to see how the terms will combine.

A check on the six d-squares now shows that each of them has been taken = 1 on at least three of the maps. Hence the restriction on unused conditions has been satisfied, and no changes need be made in Table II.

The worksheet on which the network is planned is shown in Fig. 14. The lines drawn between terms designate disjunctive combinations or factoring; and the symbols adjacent to the lines indicate which contacts are shared in each case. A careful comparison of this worksheet with the resulting network, shown in Fig. 15, will enable the reader to understand both.

Table II. A List of Selected p-Subcubes

Z	T	S	O	F
(1) 45′	(2) 45	(10) 1′2′4′5′	(6) 12	(8) 12′5
(3) 12′5′	(4) 12′5	(12) 25	(7) 12′5′	(13) 45′
(5) 1′2	(11) 25′	(14) 45	(9) 1′2′4′5	(15) 1′2′4

Three-Dimensional Maps

Up to this point, we have discussed functions of no more than four variables. If it is desired to increase the number of variables on a map, two possibilities suggest themselves:

1. Increase the number of variables plotted on each axis.
2. Use three mutually perpendicular axes instead of two.

Both methods are feasible. If method 2 is employed, then for (even) n variables, we will have $n/2$ on each axis. This means an array of $2^{n/2}$ by $2^{n/2}$ squares. However, with more than two variables on an axis, the definition of adjacence must be extended rather tenuously and subcubes become more difficult to recognize. This scheme is like the one originally suggested by Veitch.[3]

We have chosen method 2, which allows a 50-per-cent increase in the number of variables without any extension of the rules. Thus, for six variables, the methods we have described still apply, but in three dimensions.

A suitable framework is shown in Fig. 16. It consists of four 6-inch square plexiglass sheets supported at 1-½-inch intervals by rods of the same material. The rods and sheets are glued together. The author has been told that the 3-dimensional ticktacktoe boards sold at some toy shops under various names are satisfactory.

Each sheet is ruled at 1-½-inch intervals parallel to both pairs of edges. Thus we have a 4-by-4 array of squares on every sheet. The plexiglass framework enables us to do away with the writing and erasing which would be necessary when dealing with similar problems by other methods. In using it, we employ movable markers, such as 7/8-inch plastic roulette chips. The following scheme is suggested:

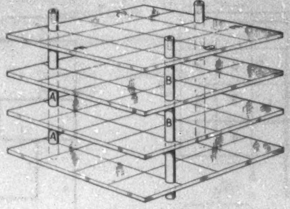


Fig. 16. The cube: a 3-dimensional plastic framework for maps

1. Mark all p-squares with white chips.
2. Mark all d-squares with black chips.
3. As subcubes are selected, mark each one with a set of distinctively colored chips.

Chips of eight or nine different colors are usually sufficient to make all the selected subcubes easily distinguishable. The corresponding products are then found by means of labels on the edges of the plastic cube.

One satisfactory labeling scheme is shown in Fig. 16. The two bottom planes are A, while the middle two are B. The variables C, D, E, and F are arranged on each plane as on the top, each letter serving to label two rows or columns. Opposite ends of any row, column, or vertical on the cube must be considered adjacent. Then every subcube may be thought of as a rectangular parallelepiped with edges 1, 2, or 4 units long. For multioutput problems, it is best to have a set of cubes, one per output.

The extension to seven variables is probably best accomplished by placing two cubes side by side. Corresponding squares in the two cubes must be considered adjacent when looking for p-subcubes. Eight variables can be handled with a set of four cubes, and nine variables require eight cubes. In the latter case, it is convenient to make them so as to stack easily into two layers of four each. Beyond nine variables, the mental gymnastics required for synthesis will, in general, be formidable. Other methods are even more limited in this respect. Outstanding exceptions to this limitation are the symmetric and positional circuits, discussed by Keister, Ritchie, and Washburn.[4]

Conclusions

Employment of the map method seems to be profitable when nontrivial problems in combinational circuit synthesis arise. Its most important advantages appear to be flexibility and speed. Further, if such problems arise frequently, it is advantageous to have a method, such as this, which can be learned and used effectively in a short time by designers new to the field.

References

1. THE DESIGN OF SWITCHING CIRCUITS (book), **William Keister, A. E. Ritchie, S. H. Washburn.** D. Van Nostrand Company, New York, N. Y., 1951, chap. 5.

2. SYNTHESIS OF ELECTRONIC COMPUTING AND CONTROL CIRCUITS (book), **Staff of the Harvard Computation Laboratory.** Harvard University Press, Cambridge, Mass., 1951, chap. 5.

3. A CHART METHOD FOR SIMPLIFYING TRUTH FUNCTIONS, **E. W. Veitch.** *Proceedings*, Association for Computing Machinery, Pittsburgh, Pa., May 2, 3, 1952.

4. See reference 1, pp. 55–64.

5. See reference 1, pp. 295–297.

◆

Discussion

S. H. Caldwell (Massachusetts Institute of Technology, Cambridge, Mass.): When Shannon published his classic paper on analyzing relay and switching circuits,[1] the engineer was given a powerful method for the solution of many problems in the field of switching circuits. Unfortunately, when one attempted to use the method, there arose a peculiar sort of frustration. Given a circuit which had been designed by the methods of trial and error prevalent at the time, it was readily possible to use Shannon's techniques to investigate alternative forms. In particular, the switching algebra could be used directly for the simplification of contact networks. But the situation was different with respect to the synthesis of a network (unless it could be described by a symmetric function). In the general case, it was necessary to resort to a word statement of the required circuit characteristics and then convert this to an algebraic statement.

For simple problems, and especially those which involved a small number of variables, no difficulty was encountered, but because of their very simplicity such problems rarely needed the algebraic approach. When problems of any magnitude were attempted, the method broke down both because of the difficulty of writing word statements and because of the difficulty of converting bulky word statements into algebraic expressions.

These difficulties were resolved by the adaptation of the logical truth table into the familiar table of combinations (see ref. 1 of the paper). This mechanism enabled the designer to state his requirements in an orderly manner, and gave him a systematic means for checking the completeness of his reasoning. Moreover, the transition from the table of combinations to an equivalent algebraic statement became almost a matter of routine, depending on individual preference for simplifying the algebraic expression by inspection of the table or by algebraic manipulation.

The arrays described by Veitch (see ref. 3 of the paper) and by Mr. Karnaugh represent further development of the table of combinations into forms which are more compact, and which also have the property of making more evident the ways in which the algebraic expression of a switching function can be simplified. Of course, the end result desired in all cases is a minimization of the required circuit, whatever we mean by the word "minimization."

The problem of manipulating functions of many variables is much like the problem the physicist had in his development of mathematical models of atomic structure. Over a period of years he succeeded in getting better and better mathematical solutions for the hydrogen atom, but none of his methods really worked when he tried to add just one more electron. Similarly, in these various methods for reducing switching functions to minimal forms we seem to be producing better and better ways for reducing functions of four variables, but we are still rather unhappy about five and six variables. The author's plastic cube for the treatment of six variables is an ingenious extension of his four-variable array, and it certainly has the reduction properties he ascribes to it. It does not, however, have the neatness of display which is a feature of the plane map; groupings of variables are not as immediately evident, and alternative groupings are even less apparent.

Mr. Karnaugh rightly points out that the search represented by this paper is in its early stages. It should be added that the need for better methods for handling the problem in more than four variables will become acute, and it is a problem worthy of the best thinking. Recent developments in the synthesis of sequential circuits show that the end result of a sequential synthesis is a combinational problem. It is a multiple-output problem in many variables, and has ramifications which will tax the best efforts of the circuit designer. Among the possibilities for meeting this problem is that of mechanizing the process involved in the map method.

Incidentally, I am not impressed by the drawbacks attributed to the Harvard Computation Laboratory minimizing chart. The large number of entries involved is no drawback in these days of cheap duplication processes. Keeping track of the entries is really a simple routine. In using the chart for the realization of six-variable functions with don't-care conditions, I find that one rarely has to complete the vertical ruling of the entire chart because the required conditions are usually satisfied with terms at the left-hand side of the chart. In some cases one finds a condition which is satisfied by only one possible minimal term, where the acceptance of that term in turn specifies the nature of one or more don't-care conditions. Of course, the six-variable cube inherently contains the same information, but it is doubtful that its display gives the designer quite as much immediate guidance as he gets from the minimizing chart.

REFERENCE

1. A SYMBOLIC ANALYSIS OF RELAY AND SWITCHING CIRCUITS, Claude E Shannon. *AIEE Transactions*, vol. 57, 1938, pp 713–23.

M. Karnaugh: In view of Professor Caldwell's remarks about mechanization, it appears to be desirable to restate the reasons for presenting this paper.

The map method, in its present form, is likely to be useful in two ways: as a pedagogic device, for the introduction of ideas about logic circuits and their synthesis, and also as a desk-top aid to the working engineer.

In making full use of the human faculty for recognizing geometric patterns at a glance, the map method supplies a number of short cuts to synthesis that are not as easily found by other methods. On the other hand, the development of machine which can recognize such relationships has only begun. If one mechanizes the map method in a more conventional way, using a repetitive scanning technique, then the result is similar to a mechanization of the Harvard minimizing charts and no special advantages are expected.

The minimizing charts, which represent one of the first significant advances over purely algebraic manipulation, have proven their usefulness in practice and will undoubtedly do so even more convincingly when machines are programmed to work along the same lines. However, it has been the author's experience that maps present the specifications for a logic circuit in a form more easily used by the human operator. Here, habit and taste enter the picture and it would be unwise to dwell on this point.

For those who are relatively new to the problem under discussion, it is suggested that a number of problems be worked by both methods. It is of interest to see how they are related, and each will throw some light on the operation of the other.

The Use of Steel Sheet for the Construction of Shielded Rooms

A. M. INTRATOR
ASSOCIATE MEMBER AIEE

LOW-LEVEL electronic or electrical measurements are particularly susceptible to errors introduced by external electromagnetic influences. The coupling of spurious electromagnetic energy into a measuring system not only may result in the receipt of false information but also can sometimes cause the complete masking of the desired data as well. For these reasons, many low-level measurements, such as the determination of crystal characteristics, filter insertion loss, noise measurements, and the like, must be made in a location as free as possible from such interference. In a laboratory, such isolation from interference is usually achieved by completely enclosing an area in copper or bronze screening. By shielding off a region relatively free of external interference in this way, a working area is provided within which sensitive electronic measurements can be made.

Certain instrumentation requires a much higher degree of freedom from extraneous influences than can be obtained in screened enclosures. A reduction in the shielding efficiency of screened booths occurs at the lower frequencies because of practical limitations in wire size and at higher frequencies because the wave lengths begin to approach the dimensions of the mesh openings. Fig. 1 shows a typical attenuation curve of a screened room. Because sheet metal presents neither of these difficulties, it is often used instead of screening, to enclose those areas in which a high degree of shielding is required.

Copper sheet has ordinarily been used for this purpose, although copper-clad steel has been used in some cases. Such rooms are usually double-walled; the inner and outer sheet-metal walls are spaced about 4 inches apart and are insulated from each other except at the point where the power line enters the room. Ordinarily the walls are hung upon a kiln-dried, wax-impregnated wood frame and all seams and mounting nails or bolts are completely soldered over to reduce the possibility of energy leakage into the room. Special seals are used to insure continuous metal-to-metal contact around the periphery of the door. Air is introduced through "wave guide below cutoff" vents, the cutoff frequency being determined by the expected top operating frequencies in the room. All power lines entering the room are filtered.

These rooms are very expensive, having ranged in cost from about $10,000 for small rooms to $100,000 for much larger

---

Paper 53-197, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted February 11, 1953; made available for printing March 30, 1953.

A. M. INTRATOR was formerly with the United States Naval Civil Engineering Research and Evaluation Laboratory, Port Hueneme, Calif., and is now with the General Electric Company, Syracuse, N.Y.

The author wishes to express his appreciation to the members of the staff of the Stanford Research Institute Engineering Division, who performed most of the work described in this paper; to Dr. D. L. Benedict for his efforts in directing the Stanford Research Institute program; and to Dr. C. R. Freberg, formerly of the United States Naval Civil Engineering Research and Evaluation Laboratory, for the suggestions which led to this investigation.

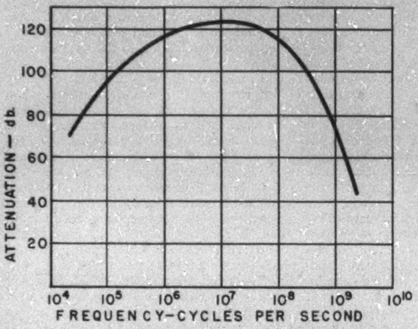


Fig. 1. Typical attenuation curve of a cellular-type double-walled screen room (22 mesh copper screening)

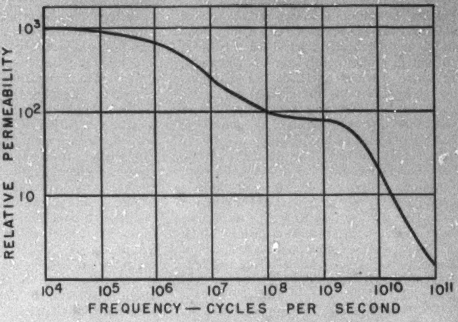


Fig. 2. Variation of the relative permeability of iron with frequency

ones. Much of the expense results from the extensive use of copper. Copper not only is an expensive metal but also requires tedious and correspondingly costly assembling techniques when being used for a shielded room. In addition, copper is at present a critical defense material in short supply. Four thousand pounds of 16-ounce copper sheet was required in the construction of one 12-foot by 16-foot by 8-foot double-walled room. For these reasons, a study was initiated to explore the possibility of using less critical and less expensive materials in the construction of high-performance rooms. A review of the theory of shielding, together with an economic comparison of the use of various metals for shielded rooms, pointed to the possible use of standard structural steel sheet.

Principles of Shielding

A shield is a physical barrier introduced between two regions to prevent the leakage of energy from one region into the other. Although no shield can be perfect, a shield can be so made that the ratio of energy levels of the two regions is very large. This shielding is accomplished in two ways: through the reflection of energy from the surface of the barrier, and through the absorption of energy within the barrier.

To illustrate this, consider the shield as an infinite plane of finite thickness and assume that the energy incident upon the shield is in the form of plane waves. In this case, for an electrically thick shield in which the absorption is of such magnitude (of the order of 10 decibels or more) that the effect of reflections at one surface does not appreciably affect the conditions at the other surface, it can be shown that the total shielding effectiveness is[1]

$$S=8.686\sigma R_s d+10\log_{10}\frac{(\eta+Z_s)^4}{16\eta^2 Z_s^2}\text{decibels} \quad (1)$$

where

the first term in the expression represents the absorption loss
the second represents the reflection loss
σ = conductivity of the shield, mhos per meter
d = thickness of the shield, meters
η = impedance of plane waves in free space, 377 ohms
Z_s = impedance of the wave in the metal = $(1+j)\sqrt{\pi f\mu/\sigma}$
f = wave frequency, cycles per second
μ = absolute permeability of the metal, henrys per meter
$R_s=\sqrt{\pi f\mu/\sigma}$
S = shielding attenuation, decibels

For nonferrous metals, the permeability is relatively independent of frequency and is equal to μ_0, the permeability of free space. However, for ferrous metals, μ is a function of frequency, as well as of the magnetic field strength and the previous magnetic history of the material. The permeability of steel decreases with increasing frequency, in a manner illustrated by Fig. 2.[2] If we assume this to be the permeability of the steel for computation purposes, Fig. 3 represents the variation of reflection loss and absorption loss with frequency of 24-gauge sheet copper and 24-gauge sheet steel.

Except at frequencies of the order of 10^{10} cycles per second or higher, the absorption loss in the steel is greater than that in copper. On the other hand, the reflection loss for the incident plane waves is higher at the copper surface than at the steel. In both cases, these theoretical attenuation values are much larger than would necessarily be required in practice. Most electronic equipment is not ordinarily sensitive to signals of field strengths of less than 0.01 microvolt per

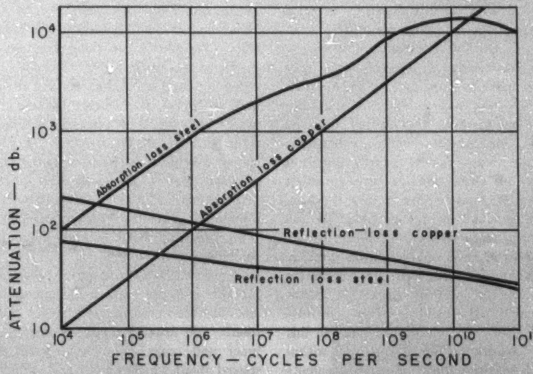


Fig. 3. Absorption and reflection loss of copper and steel as a function of frequency

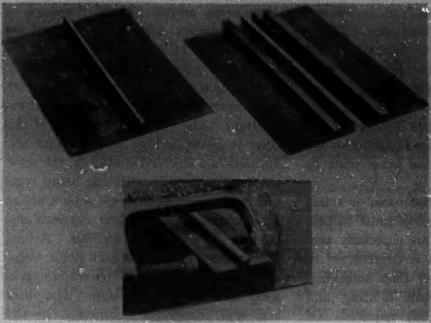


Fig. 4. Procedure used in brazing steel panels to form wall of room

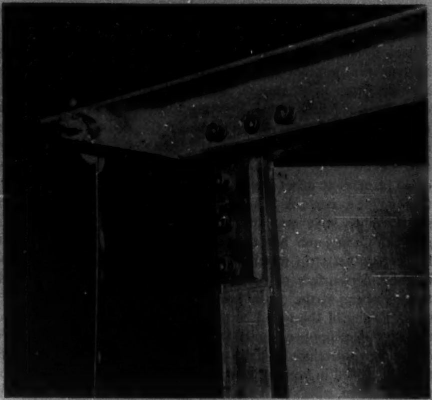


Fig. 5. Door jamb corner construction

meter. The dielectric breakdown strength of air at normal temperatures and pressures is in the order of 10^6 volts per meter for a sphere gap.[3] The maximum attenuation that could possibly be attained under these conditions is 280 decibels.

Generally, however, the shield is not a plane and the incident energy is not in form of a plane wave. Inasmuch as the reflection loss is a function of the shape of the shield and the type of impinging wave, while the absorption loss within the metal is relatively independent of these characteristics (provided the radius of curvature of the surface of the shield is ten times the skin depth), the attenuation loss within the shield is often used to determine the thickness of metal needed for a specific shielding performance. This thickness is given by

$$d = 0.114 \frac{S}{\sqrt{\pi f \mu \sigma}} \text{ meters} \qquad (2)$$

Thus it is seen, from equation 2 and Fig. 3, that the thickness of metal necessary to achieve a given attenuation decreases with increasing frequency. It would therefore appear sufficient to make calculations only at the lowest frequency in selecting the thickness of metal required to give a specific attenuation over some frequency range.

Design and Construction

The possibility of using ferrous metals for the construction of high-performance shielded rooms is clearly implied in the theory. Except at very high frequencies, the greater permeability of ferrous metals make these metals superior to copper in shielding effectiveness, for the same thickness of metal.

To verify these conclusions and also to study economical construction methods of constructing rooms made of steel, an all-steel shielded room was constructed. Because it happened to be available at the time, Tran-Cor *72*, no. 24-gauge sheet steel was used for the construction. The manufacturer's data indicated that this steel has an initial relative permeability of 1,000 at 15 kc and a resistivity of 54 microhm-centimeters. Tran-Cor *72* has the advantage of a high initial permeability with satisfactory ductility and workability for easy bending during fabrication. According to the analysis presented earlier in the paper, steel with these characteristics should offer an absorption loss of 57 decibels at 15 kc and higher attenuations at higher frequencies.

Other commercial steels can be used; and a low carbon steel, such as SAE *1010*, which is usually available from warehouse stock throughout the country, would perhaps be more satisfactory.

The size of the available sheet stock was 28 inches by 8 feet. In order to use these sheets most efficiently in building the room, an 8-foot by 8-foot by 9-foot room was designed.

The joints used in fastening the sheets together are shown in Fig. 4. The edges of the sheets were braked into 2-inch 90-degree bends; the sheets were then joined by clamping the edges together in 2- or 4-foot sections, as shown. To fuse the joint, 0.062-inch brass or Tobin bronze wire was laid in the groove, brazing flux was applied, and an acetylene torch was moved along the joint. The clamping fixture protected the panel from the flame and, by holding the joint rigidly, minimized distortion. It was found that about 6 feet of joint could be welded an hour.

Each wall, the ceiling, and the floor were assembled horizontally in this manner; the brazed units were then set up and assembled in the same way; short sections were successively clamped together and the joints brazed. For the most part, the brass wire was fed into the joint by hand as brazing progressed. Because the panels had been carefully made of sheets of standard size, no difficulty was encountered in matching sections or in the warping of sections. Because the braked edges at the joints provided enough rigidity and structural strength for a room of this size, no supporting framework was required except that necessary to compensate for the weight of the door.

The door jamb consisted of 2-inch by 1-inch steel-bar stock, into which a 1/4-inch by 3/8-inch groove was milled. A light 3/8-inch by 3/4-inch by 1/8-inch bar was spot-welded to the door jamb, as shown in Fig. 5, to provide additional support. Horizontal support for the framework of the jamb was obtained by a 3-inch by 1-inch by 1/4-inch channel and a light 1-inch by 1/2-inch angle iron was fitted along each of the corners for increased rigidity. This was securely attached to the door by brazing and bolting. Care was taken to braze and sweat solder around the bolts and to tack weld at other points.

The door was made of the same sheet steel as that used in the construction of the room, spot-welded to a welded steel frame of 1-inch by 1/2-inch steel channel reinforced by 1/4-inch by 3-inch steel spreaders. Its over-all dimensions were 3 feet by $6^1/_2$ feet. To secure a continuous firm contact between the door and the jamb upon closure, hinges and clamps were spaced at $8^1/_2$-inch intervals to provide equal pressure. Turnbuckles and braces were used at the top and bottom of the door so that adjustments could be made to offset any door warpage that might occur. A 3/8-inch diameter solid rubber strip was fitted into the groove in the jamb bar to cushion the bronze contact strip which was soldered to the inner surface of the door around the periphery at the contact area. The slight flexing of the metal as the rubber tubing was compressed upon the closing of the door introduced a little wiping action that aided in keeping the contacting surfaces clean.

During the tests, it was found that an iron-to-iron pressure connection was inadequate for shielding purposes. The contact resistance at such a closure was high enough to allow considerable energy leakage through the door closure even when the iron surfaces were newly polished. Although the single bronze strip on the door appeared adequate for the purposes of the short-term test, bronze stripping should be used on both the door and the jamb for more permanent installations.

In the experimental room, the latching mechanism could be operated only from within the room. To permit, for safety purposes, entry from the outside when the mechanism was secured, an emergency entry port was provided. This port consisted of a hole 6 inches in diameter cut near the latch handle, over which a $6^3/_8$-inch diameter cover of the same sheet steel was sweat-soldered with a soft lead solder. A tab and handle were fastened to the cover so that the solder joint could be pulled apart easily. On a more permanent installation, a simpler means of opening the door, from both inside and outside the room, is necessary. It has been the experience of the author that little energy leakage occurs at the door latch handle if the latching mechanism and handle is securely grounded to the door. Such a mechanism is shown in Fig. 6. The contact strip is fastened to the handle and rides on the brass plate as the handle is turned.

To use such a shielded room as a test chamber, auxiliary services must be provided. Such services are especially necessary when the room is constructed of solid sheet walls rather than screening, because with such walls lighting and ventilation are absolutely required. Of course, these services must be provided in such a manner that they decrease in no great way the over-all shielding performance of the room.

Ventilation

Ventilation does not present any serious problems. The ventilator can consist of "wave guide below cutoff" ducts arrayed in a honeycomb fashion to form the inlet or outlet port. The maximum cross-sectional dimension of the individual ducts should be smaller than 1/2 wave length of the highest frequency to be shielded against. Automobile radiators have often been satisfactorily used for this purpose.

Power and Lighting

The major energy leakage into shielded rooms is usually that resulting from interference carried on the power lines; for this reason, it is necessary to filter the incoming electric power properly. Each line should be individually filtered. The basic circuit of the filter designed for use with the test room is shown in Fig. 7. To improve the performance of the fiber at higher frequencies, additional sections were added. These consisted of four inductors, each composed of six turns of no. 14 wire wound on a 1-inch arbor and spaced six turns to the inch. The inductors were by-passed by 1,000-micromicrofarad "feed-through" type capacitors. The performance curve of this filter is given in Fig. 8. The filter was encased in a shielded container, the separate sections of the network being compartmented by sheet-metal separators to eliminate mutual coupling between one part of the circuit and another. A bare copper wire, soldered at each point where it passed through the separators, was used as a common ground. The filter box was connected electrically to the shielded room at only one point through a section of steel tubing soldered tightly to box and the wall. This precaution was taken to minimize the circulating currents in the room's walls that would arise if multiple contact was made. A number of filters are now available commercially that have been specifically designed to be used with shielded rooms.

The filtered power lines entering the room should be enclosed in rigid steel conduit, the conduit being soldered or brazed to the wall at intervals not greater than 1/4 wave length of the highest frequency at which the room is to be used.

Illumination is provided only by incandescent lamps since fluorescent lights have been found to be a serious source of radio interference.

Test

The problem of obtaining reliable data on the performance of high-attenuation shielded rooms has been a difficult one. Ideally, the room should be isolated in space and immersed in a uniform high-intensity electromagnetic field of known magnitude. The field strength should then be measured and integrated over the interior of the room by sensing devices and instruments too small to perturb the distribution. The reduction in the strength of the field as measured throughout the volume of the room would be an indication of the shielding effectiveness of the room.

However, in the practical case, shielded rooms are used under a large variety of circumstances but never in the ideal one. They are situated near masses of dielectric and conducting media which give rise to many secondary electromagnetic effects that disturb the fields to which the room is subjected. Further, these fields are rarely plane and include low-impedance distributions as well as high-impedance ones.

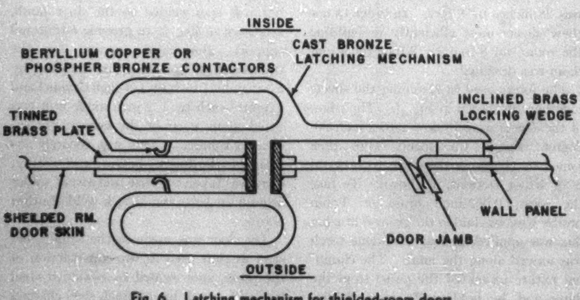


Fig. 6. Latching mechanism for shielded-room doors

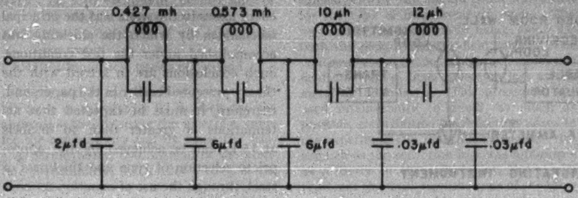


Fig. 7. Shielded-room power-line filter circuit

These limitations have generally been recognized by engineers and consequently many qualitative, though practical, methods have been devised for determining the adequacy of the particular shield under the conditions in which it will be used. Such methods usually consist of placing a transmitter within the room and a sensitive receiver outside of the room and then measuring the variation in attenuation of the transmitter's field when the door to the shielded room is opened and closed. Most shielded-room evaluation has been made in this manner although for more abstract studies some have put rooms on trucks and taken them to regions of high-intensity fields, such as those near a transmitter tower, in order to measure the degree of shielding achieved by the room.

Inasmuch as the primary objective of this study was to verify those theoretical observations that pointed to the use of steel for the construction of shielded rooms it was felt that a practical test method that would allow the performance of the room to be evaluated as quantitatively as possible in its operating location should be sufficient. It was decided to produce the test field by a source within the room and a magnetic dipole radiator was chosen for this purpose. Magnetic radiators offer two main advantages: their fields in the near region are of the low-impedance type most effective in coupling to a low-impedance conducting sheet, such as steel, and magnetic loops are easy to monitor. An ammeter can measure loop current. Knowledge of the current in the loop and its relative position in the room is sufficient to calibrate and reproduce its performance.

It was realized that this technique might be difficult to use at very high frequencies because the increased driving impedance of the loop would demand an extremely high-powered transmitter to reproduce the loop currents and corresponding fields used at the lower test frequencies. Limitations in the funds available prevented further exploration at these higher frequencies. However, the results obtained with the loop radiator can be reasonably expected to give the minimum effectiveness of the steel as a shielding medium, inasmuch as the difficulty with which the energy penetrates the shield decreases as the frequency is lowered. These measurements, therefore, should check the performance under theoretically the most stringent conditions.

Measurements were made in two frequency regions, 2 to 15 megacycles, and 10 to 30 kc. At the maximum frequency of 15 megacycles, the transmitter loop consisted of a copper rod 3/8 inch in diameter, formed into a single-turn 20-inch diameter circle and having a measured inductance of 1.23 microhenrys and an unloaded Q at 10 megacycles of 255. A 0 to 15-ampere radio-frequency ammeter was used to measure the loop current and a 150-watt transmitter was used to drive the loop.

Within the 2-to-15-megacycle range, a standard high-sensitivity radio receiver utilizing an unresonated receiver loop identical to the transmitting loop was used as the field detecting device. Because the built-in shielding of the receiver was inadequate, the receiver was always operated inside the shielded room.

The general test procedure involved substitution techniques. The transmitting and receiving loops were mounted on a table coaxially and spaced 8 inches apart. The transmitting loop was fed by a short, parallel-wire transmission line, and the receiving loop was connected to the receiver through a 6-foot coaxial cable. A calibrated 170-decibel variable attenuator was placed in the receiver transmission line. The receiver which was located inside the shielded room, was connected to the loop through a coaxial fitting in the wall. A 10-decibel attenuating pad was placed between the receiver and the loop transmission line to isolate the loop from the receiver. This pad was within the room, but the larger range of attenuators was left outside. There was no evidence of radiated leakage into the attenuators.

The transmitter was then adjusted to maintain almost 4 amperes of loop current. At this excitation, nearly the full 170 decibels were required of the receiving loop attenuator to give a low sensitive voltage indication on the automatic volume control of the receiver. Readings were taken of the transmitting loop current and the automatic volume control voltage. The receiver loop was then placed inside the room, the transmitting loop being kept outside. The loops were moved to coaxial positions, parallel to and equidistant from the wall of the room and separated by the same total distance previously used. The transmitting loop current was adjusted to 4 amperes, and the receiving-loop attenuators were adjusted so that the receiver indicated the same automatic volume control voltage as that used during the calibration procedure. The difference in the attenuator settings required to obtain the same

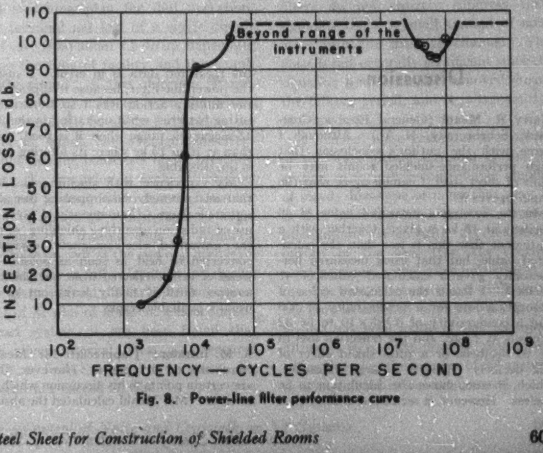
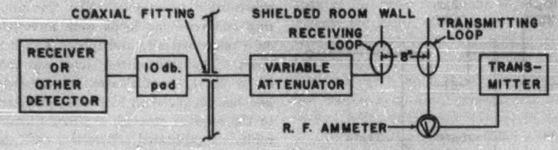


Fig. 8. Power-line filter performance curve



a) ARRANGEMENT FOR CALIBRATING INSTRUMENT

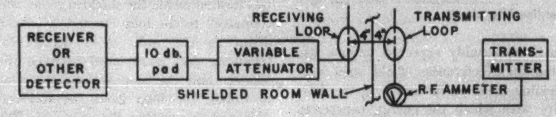


Fig. 9. Block diagram of test setup used for evaluating the performance of the shielded room

receiver automatic volume control voltage in the two setups was taken as the attenuation introduced by the room. A block diagram of the test procedure is shown in Fig. 9.

Because it was possible that some of the received energy may have been introduced through the power-line filter, a battery power source was installed inside the room and the measurements were repeated. No change was noticed.

A similar substitution procedure was used to obtain the attenuation at 15 kc. For this frequency, a receiving loop with a diameter of 20 inches was made up of 100 turns of no. 28 enameled copper wire. This loop had a measured inductance of 13.38 millihenrys and an unloaded Q of 33 at 15 kc. A harmonic analyzer was used to measure the loop voltage.

Results

Employing this method, the measured attenuation of the room was found to be 46 decibels at 15 kc and greater than 160 decibels between 1 and 10 megacycles. These values are somewhat less than those which had been calculated for the thickness of steel used. However, the calculated results were based upon the value of the permeability of the steel which had been furnished by the supplier—a relative permeability of 1,000 at 15 kc. Subsequent measurements of the permeability indicated that the actual value was closer to 500. When the corrected value is used in the equation, a calculated attenuation of 41 decibels is obtained, which value agrees fairly closely with the experimental results. The permeability difference may have resulted from the mechanical working of the metal when it was being assembled.

Checks made of the attenuation between 10 and 30 kc showed that the measured shielding in decibels was proportional to the square root of the frequency. This fact indicates that the attenuation through the metal evidenced by skin-effect losses in the walls was the principal mechanism by which the shielding was accomplished under the test conditions. Such conclusions are in accord with the theory presented earlier in the paper, and, therefore, it must be expected that attenuations of greater than 46 decibels at 15 kc can be achieved by the appropriate selection of type and thickness of steel, through the use of equation 2.

The test setup employed represents the most stringent conditions that might be encountered in practice: that of shielding against the near field of a partially or completely closed loop carrying a large current. The near field of a current loop is a low-impedance field. Such a field undergoes little reflection loss at metal-air boundaries; and consequently, the results represent the minimum attenuations that could be secured under the worst conditions.

Conclusions

High-performance shielded rooms may be built of sheet steel. By the appropriate selection of the thickness and type of steel, a single-walled room may be built to achieve the degree of shielding required. The modified skin-depth equation as given by equation 2, serves as a good criterion for selecting the proper type and thickness of steel.

References

1. THE ELECTROMAGNETIC THEORY OF COAXIAL TRANSMISSION LINES AND CYLINDRICAL SHIELDS, **S. A. Schelkunoff.** *Bell System Technical Journal*, New York, N. Y., vol. 13, 1935, p. 569.

2. INTERNATIONAL CRITICAL TABLES, VOL. 6 (book). McGraw-Hill Company, New York, N. Y., 1929, p. 375.

3. ELECTRICAL ENGINEERS HANDBOOK, VOL. 1 (book), **H. Pender, Wm. A. Del Mar.** John Wiley and Sons, New York, N. Y., 4th ed., 1949, pp. 5–66.

◆

Discussion

Harry R. Meahl (General Electric Company, Schenectady, N. Y.): Although I agree with the author's conclusion that high performance-shielded rooms may be built of sheet steel, I cannot agree that the paper proves it.

In the results, a measured value of 46 decibels at 15 kc is given, together with a statement that this is less than the calculated value, but that using measured permeability gave a calculated value of 41 decibels. I found the calculated value of absorption loss for a permeability of 500 and a conductivity of 0.185×10 to be 32 decibels at 15 kc, and the reflection loss to be 90 decibels, or a total shield effect of 122 decibels. This is a large difference, which, if true, shows the calculation to be useless. However, it seems more likely that the measured data is in error. I suspect the power line filter, because it appears that the filtering action was tested by substituting batteries when operating in the 2-to-15-megacycle range where it is good, rather than at 10 or 15 kc where its filtering action is questionable.

My experience with shielding has been that an extremely small opening can allow serious leakage. Consequently, I favor the use of radiation absorbing shielding, so that the unwanted electromagnetic energy is converted to heat as soon as possible instead of being reflected from one shield to another until it finally leaks out of the usually available cracks.

A. M. Intrator: I appreciate Mr. Meahl's comments on my paper. However, there are certain points in his discussion which are in error. Mr. Meahl calculated the absorption loss in the steel to be 32 decibels using a conductivity of 0.185 × 10. I presume this is in units of mho per meter. If so, since the resistivity of the steel was 54 microhm-centimeters, Mr. Meahl's value is incorrect. The correct conductivity is 1.85×10^6 mho per meter. Substitution of this value into equation 2, and using a relative permeability of 500 for the steel still results in a calculated absorption loss of 41 decibels, as shown

$$\text{Absorption loss} = \frac{0.635\times10^{-3}}{0.114}\times \sqrt{\pi(500\times1.257\times10^{-6})(1.85\times10^{6})}\times (1.5\times10^{4}) = 41 \text{ decibels}$$

where

0.635×10^{-3} = thickness of 24-gauge sheet steel, meters

$500\times1.257\times10^{-6}$ = absolute permeability of the steel, henrys per meter

1.85×10^{6} = conductivity of the steel, mho per meter

1.5×10^{4} = frequency, cycles per second

I believe that Mr. Meahl may have made his error in calculating 32 decibels for the absorption loss by using the incorrect value for the conductivity and by using the relative permeability, 500, instead of the absolute permeability, $500\times1.257\times10^{-6}$, in the equation.

Consideration of the reflection loss in the calculations by Mr. Meahl is not warranted for the type of measurements made. The measurements were made using loops for both the pickup and the field source. The field near a current loop is a low impedance field and suffers little reflection loss at the metal-air boundaries. This procedure was used intentionally to avoid, as much as possible, contribution by reflection losses to the experimentally determined results. Inasmuch as the reflection loss was minimized as a factor in the experiment, the results represent the most rigorous type of shielding performance obtainable, since little or no additional shielding was provided by reflection losses.

The tests to determine whether or not some of the received energy was coming in through the power line filter, by substituting batteries for the power source, was made at 15 kc as well as in the 2-to-15-megacycle range.

I fully agree with Mr. Meahl that extremely small openings can cause severe leakage. However, the purpose of this paper was to point out that steel can be used as a substitute for copper in constructing high-performance shielded rooms.

The Control of Noise and Crosstalk on N1 Carrier Systems

A. J. AIKENS
NONMEMBER AIEE

C. S. THAELER
ASSOCIATE MEMBER AIEE

The 12-channel *N1* carrier system is designed primarily for short- and medium-haul service and to make use of existing, and new, cable plant with a minimum of special treatment. As compared to other types of carriers, the lower transmission levels of the *N1* system tend to make it more susceptive to interference from external sources. The use of higher line frequencies with the corresponding decrease in coupling loss between line conductors tends to increase the amount of crosstalk between carrier systems. While the technical features of the *N1* system have already been described,[1] it may be well to review some of the characteristics that are of interest in connection with the control of noise and crosstalk on message circuits. These include:

1. Transmission of both side bands and the carrier.
2. A compandor per channel. At the transmitting end, a compressor raises the average talker volume by about 10 decibels (db) and reduces by one-half the normal range of talker volumes. At the receiving end, a complementary expandor restores the talker volumes to their normal range.
3. Transmission in the two directions in different frequency ranges on separate pairs in the same cable. Low-group frequency range, 44 to 140 kc; high-group range, 164 to 260 kc; frequency inversion at each repeater by modulating the signal with a 304-kc carrier.
4. In each repeater section the carriers at the high-frequency end of the band are transmitted at higher level than those at the low-frequency end. Because of the increase in attenuation with frequency, the relative levels at the receiving end are reversed.

Control of Noise

Noise Objective

In general, the effective noise on an *N1* channel should not exceed 31 decibels adjusted (dba) referred to the zero level point. Under the Bell System general toll switching plan for dial operation, the terminal net loss of a single circuit will vary from about 4.5 db up, depending upon the type of facility and the length of the circuit. If this objective is met, the maximum noise on even the lowest loss circuit will be about 5 db below the level at which the average subscriber would become conscious of its presence. Since the expandor in each *N1* channel terminal is designed to have a hangover time of about 40 milliseconds to avoid clipping the trailing ends of syllables, the noise present at the end of a high level signal is not as fully attenuated as during the silent interval. To achieve the equivalent of 31 dba of noise on a noncompandored system, it has been found that the silent interval noise, as measured with a Bell System 2-type noise set, should not exceed 26 dba.

Sources of Noise

The principal sources of noise encountered in the operation of *N1* systems are as follows:

1. System noise.
2. Secondary induction from telephone plant.
3. Secondary induction from external sources.
4. Nonsoldered cable pair joints.

Discussion of Noise Sources and Mitigation

System Noise

By design, the noise contributed by the repeaters and terminal equipment has been held to about −38 dba of equivalent input noise for a 200-mile circuit. This noise consists of the usual first-circuit *sh-sh* (thermal-type noise) and interchannel modulation due to the nonlinearity of the modulators and demodulators. The noise from these sources is so low as to be masked by noise from other sources.

The *N1* system uses "out-of-band" signaling employing both side bands of a 3,700-cycle tone, transmitted 15 db below the carrier level. Since the channel filters do not entirely suppress these signaling tones, a small amount of noise leaks into the message channel. This noise is not generally significant since the signaling tone is usually removed during the talking period, and it is transmitted at a low level.

A thermal noise generator is provided in the transmitting terminal of the *N1* system to furnish a controllable amount of steady hiss-type of noise that may be used to mask low level crosstalk or crossmodulation. Even with the maximum output of the generator, the thermal noise produces a negligible impairment to message circuit transmission while masking the crosstalk by about 5 db. So far, no use has been made of this technique although it is quite possible that it will be employed as the number of *N1* systems in nonquadded cables is increased.

Paper 53-226, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 18, 1953; made available for printing May 14, 1953.

A. J. Aikens is with the Bell Telephone Laboratories, Inc., New York, N. Y., and C. S. Thaeler is with the American Telephone and Telegraph Company, New York, N. Y.

The authors wish to acknowledge the efforts of the many persons who made this study possible, especially the work of S. E. Church, S. G. Hale, H. A. Stone, Jr., R. S. Tucker, and M. A. Weaver, all of Bell Telephone Laboratories.

Secondary Induction from Telephone Plant

The operation of relays and switches in a telephone office puts noise voltages between the office wiring and ground. While the office wiring associated with the *N1* systems consists of shielded pairs, this noise is picked up by noncarrier wiring over which it travels longitudinally to pairs in the same cable sheath as the *N1* pairs. Due to the high longitudinal coupling between all pairs under the same sheath, noise voltages appear on the *N1* pairs, and depending upon their longitudinal-to-metallic balance, part of the noise appears in the metallic circuit. When installing *N* carrier on existing cables, it is not generally feasible to improve the cable balance although it is possible to exclude from *N* operation pairs having particularly poor balance. The path by which noise from internal sources enters the *N* system is shown in Fig. 1.

Since little can be done to reduce the noise after it has entered the *N* cable, it is desirable to suppress it before it reaches the carrier pairs. This might be done by introducing a high impedance into the longitudinal circuit of the noncarrier pairs without appreciably increasing their metallic circuit loss in the useful frequency range. With *K* carrier systems this was accomplished by means of series retardation coils in the noncarrier circuits. This treatment proved to be expensive, and a cheaper way of suppressing the noise on *N* carrier systems is still being sought.

One measure that has been used is to keep the carrier and noncarrier pairs in separate sheaths or shielded compartments for a limited distance from the office. As the length of the twin cable section is increased, the received signal level on the *N1* pairs at the point of exposure to the noise is increased and the noise itself is reduced, thus greatly improving the signal-to-noise ratio.

Noise arising from the operation of telegraph equipment may be effectively suppressed at the source. In many cases the telegraph equipment already includes noise killers to control the noise in the audio-frequency range. These may be supplemented by suppression that is more effective at carrier frequencies. Not all types of telegraph systems cause serious interference. With metallic operation, the currents are so small as to create negligible disturbance. However, the large currents used with grounded simplex operation have proved bothersome. Experience has shown that a simple filter consisting of a 0.5-microfarad capacitor connected between ground and the mid-point of the simplex-deriving repeating coil will provide adequate suppression.

One of the more commonly used measures to control dial office noise is to limit the length of the repeater section adjacent to the noise source and thus restrict the amount of noise gain. The amount of metallic noise on a cable pair at the main distributing frame in a dial office will vary widely but it has been found under typical conditions to be in the order of 100 decibels below 1 milliwatt in a 3-kc band in the *N1* frequency range. This compares with a side-band level at the input of a high-low repeater which may drop to 75 decibels below 1 milliwatt for a full 8-mile repeater section. At a high-low repeater the 304-kc carrier modulates noise components in the frequency range between 348 and 444 kc, the lower side band of this modulation also falling within the 44- to 140-kc low-frequency group. This image frequency noise is excluded from low-high repeaters and low-group receiving terminals by the low-pass filter, and is no problem at high-group receiving terminals since the 304-kc group carrier is not present.

Secondary Induction from External Sources

The most important path for induction from sources external to the telephone plant is by secondary induction from noncarrier pairs in the *N1* cable which have extensions or taps to open-wire lines or to branch cables of short length connected to open wires or subscriber drops. Noise from atmospheric static or radio-frequency transmitters which is present on the open wire or drops is carried into the cable longitudinally where it appears as metallic noise on the *N1* pairs, as previously discussed. Several different measures have been used to suppress this type of noise, depending upon its magnitude, frequency, and nature.

Where suppression is desired over a wide range extending through the transmitted *N1* frequencies and the image frequency band, the most effective suppression is provided by longitudinal retardation coils which insert loss in the ground return circuit. Existing coils which were developed for use with the *K* and *J* carrier systems have been used to a limited extent but they are expensive and do not have adequate loss in the image band. New coils are under development and some prototypes have been tested with satisfactory results. It appears that they will afford about 25 db of suppression throughout the entire *N1* transmitted and image frequency bands. In developing the new coil every effort is being made to keep its metallic circuit loss at a minimum. Fig. 2 shows a longitudinal retardation coil connected at the junction of open wire and cable.

The experience with *N1* systems in the vicinity of Annapolis, Md., will serve to illustrate the application of this type of suppression. On many systems in this area, channels 3 and 7 were turned down because of interference from nearby radio transmitters. The noise was introduced into vulnerable points of the *N1* circuits by way of open-wire and drop-wire extensions of non-*N* cable pairs. The interference had been anticipated when engineering the carrier systems, and as a result the two or three repeater sections nearest the radio transmitters were reduced in length to facilitate its suppression.

After placing the initial systems in service, experiments were begun to see what could be done to reduce to tolerable levels the noise on the affected channels. Between Annapolis and Hyattsville there were 354 different open-wire and drop-wire extensions of non-*N* pairs, of which 106 required treatment to acceptably reduce the *N1* channel noise. The only major group of extensions requiring no suppression were in the $3\frac{1}{2}$-mile-long repeater section at the Annapolis end of the system. The suppression finally

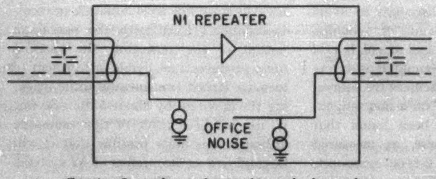


Fig. 1. Secondary induction from telephone plant

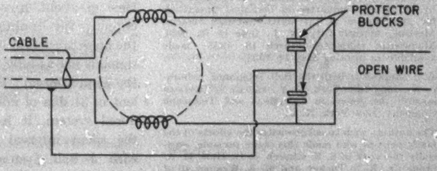


Fig. 2. Longitudinal retardation coil

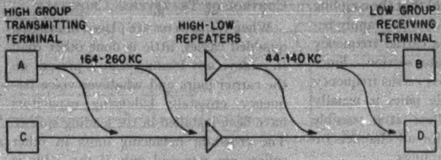


Fig. 3. Transverse far-end crosstalk

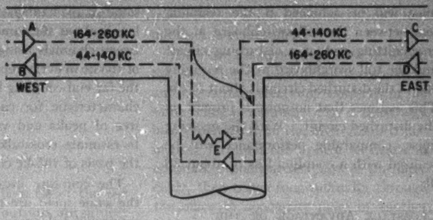


Fig. 4. Crosstalk in lateral cables

selected consisted of toroidal coils such as shown in Fig. 2, wound so as to insert a relatively high impedance in the longitudinal circuit of the offending antenna-like pairs but not in the metallic circuit of these pairs. A design including high core loss gave a large resistive component to the impedance to aid in preventing resonance.

Table I shows how the noise in the affected channels was reduced by the installation of the 106 separate suppressors.

An interesting aspect of the tests was that subsequent removal of suppression from a single open-wire extension 1,500 feet long, connected to the cable at a point about 12 miles from the radio transmitters, raised the noise nearly to the level existing before any extensions were suppressed.

In some cases noise has been experienced on systems operating on cables without open-wire taps or drop loops. Here the longitudinal noise currents present on the sheath encountered a high impedance due to the presence of an insulating joint and the noise was thereby transferred to the longitudinal circuit of the pairs. Certain capacitors used to by-pass insulating joints in voice frequency cables do not effectively by-pass carrier frequency currents. Where the need for the insulating joint no longer exists, it has been permanently removed.

Table I. Reduction of Radio Interference into N1 Systems

Channel	Noise at Zero Transmission Level at Annapolis	
	Without Suppression	With Suppression
Worst No. 3	67 dba	27 dba
Average No. 3	50	23
Worst No. 7	59	28
Average No. 7	50	22
	Noise at Zero Transmission Level at Hyattsville or Washington	
	Without Suppression	With Suppression
Worst No. 3	68	30
Average No. 3	52	27
Worst No. 7	49	25
Average No. 7	37	21

In other cases the joint has also been bridged by a 4-microfarad capacitor.

Some use has been made of slot rejection filters as a stopgap measure until the installation of suppression which does not impair the *N1* message circuit. To minimize the impairment to speech transmission, only narrow-band slot filters are employed and this means, of course, that the method is satisfactory only when the interference is a single frequency tone. It is admittedly not a desirable measure since the effectiveness of the suppression is reduced by variations in the interfering frequency. It is obviously undesirable to use slot rejection filters on circuits which may be operated in tandem with other circuits which may also contain filters tuned to a different frequency.

Nonsoldered Cable Pair Joints

Occasionally *N1* systems are operated in part or entirely over cable pairs with nonsoldered joints. This usually occurs only in the case of older exchange cables since it is now common practice to solder new cables which are candidates for *N1* operation. The defect in nonsoldered joints is due primarily to a barrier of oxides and possibly sulfides. The joint resistance may be variable, because of temperature changes or vibration, and may produce a noise sounding like thermal noise with superimposed irregular clicks and pops. Where a cable on which *N1* carrier is to be installed is known to have a bad history from the standpoint of variable resistance, it is possible to make a series of tests over a period of time to determine which pairs are properly suitable for carrier use.

The presence of d-c supplying repeaters with power is thought to help minimize noise production in the nonsoldered joints. Where such power is not supplied over the cable pairs, it is possible to supply up to 20 mils of bleeder current to aid in stabilizing the joint resistance.

Control of Crosstalk

The solution to a number of crosstalk problems present in other open-wire and cable carrier systems has been facilitated by the design of the *N1* system. For example, frequency frogging has solved the runaround crosstalk problem at repeaters without requiring the introduction of loss in the longitudinal path as in open-wire systems, or the use of separate cables which permit pair frogging as in the case of type-*K* cable carrier. Also, the compandor per channel helps to control far-end and other types of crosstalk without the use of frequency staggering as applied in 3- and 12-channel open-wire systems or the provision of the intersystem crosstalk balancing associated with type-*K* carrier.

It is possible to operate a large number of *N1* carrier systems in most of the cables now in telephone plant. In quadded cables it is expected that carrier may be operated on substantially every quad. In nonquadded cables, where the coupling losses are somewhat lower, crosstalk poses more of a problem. Anticipating this, the thermal noise generator was provided in each transmitting group unit. However, as previously noted, it has not yet been necessary to introduce this crosstalk masking noise.

CROSSTALK DESIGN OBJECTIVE

The Bell System has developed a system for rating performance as controlled by crosstalk which takes into account the distribution of talker volumes at the originating toll switchboard, the number of disturbing circuits, the rms value of the current ratios in each group of similar crosstalk couplings, and the masking effect of various types of noise, including babble or the crosstalk from multiple disturbers. The crosstalk couplings between the combinations of quads or pairs in a cable vary over a considerable range. It is evident therefore that, if a limited number of the conductors in a cable is employed for carrier operation, the probability of hearing crosstalk from other systems is reduced. Without elaborating further on the fine points of crosstalk index, as this criterion of performance is known, we may note that good performance will be achieved if the coupling loss between the toll switchboard at the transmitting end of a disturbing circuit and the toll switchboard at the receiving end of the disturbed circuit is about 35 db. This assumes that no noise is present on the disturbed circuit. With the masking noise, comparable performance will be realized with a coupling loss of about 30 db.

CROSSTALK ADVANTAGE OF THE COMPANDOR

The expandor at the receiving terminal of the *N1* channel introduces a loss of 28 db as long as the total side-band power at its input is below the level at which the expandor begins to open up. This is always true in the case of the crosstalk encountered in commercial practice. The 28-db loss effectively attenuates the crosstalk during the silent interval when it is most objectionable. If the disturbing circuit is not compandored we may take full credit for this attenuation, but if it is equipped with a compandor we must recognize the effect of its compressor on raising the average talker volume at the transmitting terminal. Under present-day conditions, the *N1* compressor increases the crosstalk potentialities by about 7 db and hence the net compandor advantage for crosstalk between *N1* systems is 28–7, or 21 db.

TRANSVERSE FAR-END CROSSTALK

Fig. 3 shows the path of transverse crosstalk in which *AB* is the disturbing circuit and *CD* is the disturbed circuit. Since the crosstalk loss per unit length varies inversely with frequency, it is evident that the coupling loss between *AB* and *CD* will be less in the first repeater section than in the second. It may also be noted that in the first section the crosstalk loss for channel 12 using a 256-kc carrier will be less than for channel 1, which uses a 168-kc carrier. However, in the second repeater section channel 12 operates on a 48-kc carrier at which frequency the coupling loss is greater than for channel 1 with its carrier at 136 kc. It has been found that with an even number of repeater sections the amount of far-end crosstalk is substantially the same on all channels and that the coupling loss is about the same as would apply between channels using a carrier frequency of 165 kc in every repeater section. Since the far-end coupling loss versus frequency characteristic for cable pairs is usually free of peaks and valleys, it is possible to estimate crosstalk for all channels on the basis of 165-kc couplings.

The coupling losses between pairs in the same quad are relatively low unless intraquad balancing is employed. However, by utilizing the two pairs of a quad for the same *N1* system, the effect of this low coupling loss is not significant since different bands of frequencies are transmitted over the two pairs. The coupling loss between equal level points on the pairs in different quads is considerably higher, the rms loss being in the order of 75 db per mile at 165 kc, and the 1 per cent minimum roughly 10 db less. These losses may vary ±5 db depending upon the size of cable and type of construction, (that is, long-pair twist or short-pair twist) but are typical of present-day cables in the 30-quad size. With couplings such as this, the transverse crosstalk in quadded cables is not serious even on systems up to 200 miles or more in length, and no cable balancing is required. When placing new quadded cables primarily for *N1* carrier use certain splicing precautions may pay dividends in lower crosstalk, particularly when the cable is of small size. This effect of splicing will be discussed later.

The situation with nonquadded cables is different, particularly in the case of older types in which the same length of pair twist was used for all pairs in the cable. For example, the rms losses in nonstaggered twist cables run about 55 db per mile at 165 kc. With improved manufacturing methods and more rates of twist, it has been possible to raise the rms coupling loss in a present-day 25- or 26-pair unit of pulp-insulated 19-gauge cable to about 68 db per mile at 165 kc with the rms coupling loss between units about 9 db higher. Even better performance has been achieved in some experimental cables utilizing polyethylene insulation and ten different rates of pair twist.

CONTROL OF TRANSVERSE CROSSTALK

When *N1* systems are placed in existing quadded cables, little is done other than to remove any loading that is present on the carrier pairs and whatever voice frequency crosstalk balancing capacitors have been installed in the loading splices. The crosstalk balancing units in other splices are removed only if the cable is opened for some other reason. Building-out stubs which were placed to provide uniform loading section length are removed since these no longer serve a useful purpose and may introduce considerable crosstalk.

When new quadded cables are placed for *N1* use they are frequently of small size since even a 27-quad cable is capable of providing 300 message channels if fully developed for carrier use. As the size of a cable is reduced, it is obvious that the chance for any two quads to occupy adjacent positions is increased. The coupling between adjacent quads is much greater than between alternate or more widely separated quads so the rms coupling loss decreases with size. To achieve the least crosstalk, therefore, it is desirable to splice the cable so as to equalize as nearly as possible the number of adjacencies between all *N1* carrier quads. Rather than leave the splicing plan entirely in the hands of chance, splices in small cables are sometimes planned using the so-called number-splicing technique. This procedure insures as thorough a mixing of the quads at each splice as is theoretically possible. With larger cables, simplified random splicing techniques are generally adequate.

More care is desirable when engineering carrier on nonquadded cables, since the over-all coupling losses are somewhat lower than in quadded cables and the couplings vary over a wider range. If the cables are in place, the amount of crosstalk is estimated, taking into account the length and coupling loss in each cable section and the number of possible disturbing channels. Should the expected performance fail to meet the desired objective, the use of cables with higher coupling losses is the first consideration. If this is not feasible, it is possible to select pairs in each repeater section so as

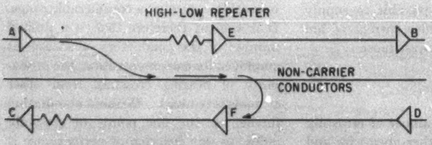


Fig. 5. Interaction crosstalk at high-low repeater

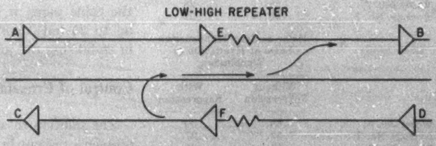


Fig. 6. Interaction crosstalk at low-high repeater

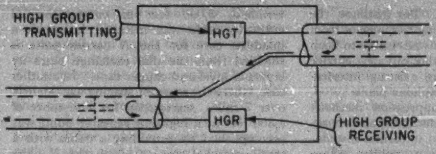


Fig. 7. Interaction crosstalk at terminal

to minimize the effects of bad coupling. This is accomplished by assigning to the same frequency group only those pairs between which the coupling losses are relatively high. In some cables there may be a few bad actors which do not lend themselves even to this treatment, and they must therefore be discarded from carrier use. The possibilities for improving performance by pair selection are great but so are the costs of the coupling measurements. Practical application of the technique is therefore expected to be rather limited.

When placing new nonquadded cables, the approach is similar to that for existing cables except for the opportunity to increase the coupling loss by controlling the splicing.

CROSSTALK IN LATERAL CABLES

Fig. 4 illustrates a situation where it has been desirable to locate a high-low repeater at the end of a lateral cable. The repeater section to the west has been assumed to require the use of a pad at the input to the repeater at *E*. The repeater section to the east requires no such pad and there is, therefore, in the lateral cable a level difference between pairs in the same frequency group. If the level difference at 256 kc is greater than about 27 db with a lateral cable length of about 200 feet, the resulting near-end crosstalk from *A* to *B* may add significantly to the crosstalk already present at terminal *B*. The remedy is to reduce the level difference in the lateral cable either by equalizing the length of the repeater sections or by moving part or all of the span pad from the input of the repeater at *E* to the output of the repeater at *A*.

As the length of the branch cable is increased, the coupling loss is decreased and the corresponding level difference at 256 kc must be reduced. When more than one lateral cable is used on the same system, it is necessary to evaluate the effects of all such cables in contributing to the near-end crosstalk.

Consideration of lateral cable crosstalk between low-frequency groups is not generally required since the coupling losses in the cable are higher. At low-high repeaters, however, differences in high-frequency levels because of the use of pads at the repeater outputs are subject to the same restrictions as just outlined.

INTERACTION CROSSTALK AT A REPEATER

Fig. 5 shows the interaction crosstalk path at a repeater by way of paralleling noncarrier conductors. If the length of the repeater section from *A* to *E* is shorter than from *D* to *F*, requiring the use of span pads as shown, there will be a level difference between channels operating at the same carrier frequency. When the 256-kc level difference at a single repeater exceeds about 27 db, the amount of crosstalk through the interaction path represented by the arrows may contribute significantly to the crosstalk at *C*. Again, since the loss in this coupling path is higher at the low frequencies, the interaction crosstalk is not generally a problem at low-high repeaters.

This type of interaction crosstalk may occur wherever successive repeater sections differ in length. It is necessary, therefore, to be sure that the combined crosstalk from all high-low repeaters does not become excessive. In general this will be the case if the 256-kc level difference between the inputs from east and west at each high-low repeater is limited to 16 db or less.

Occasionally it is desired to place a span pad or artificial line at the output of a low-high repeater as shown at *E* in Fig. 6. Since the levels at the repeater outputs are the same, the level on the cable pair at the output of the repeater at *F* will be higher than on the pair at the output of the pad at *E*. Again there is an interaction path through the noncarrier pairs, and the same 27-db level difference objective applies.

INTERACTION CROSSTALK AT TERMINALS

At the beginning of the discussion of crosstalk, reference was made to the frequency frogging at each repeater, which facilitates the control of crosstalk between high-level and low-level pairs in the same cable. In most cases, this frequency segregation of high- and low-level signals is also maintained at the carrier system terminals by specifying either "high-group transmitting—low-group receiving" terminals or "low-group transmitting—high-group receiving" terminals in a given office.

There are instances, however, in which the operation of high-group transmitting and high-group receiving terminals in the same office is desired. Fig. 7 illustrates how crosstalk from the high-group transmitting terminal can appear as near-end crosstalk on noncarrier pairs in the same cable, then travel longitudinally through the office to noncarrier pairs in another cable containing *N1* systems using high-group receiving terminals. When there is no equipment on the noncarrier pairs the longitudinal loss of the office wiring may be substantially 0 db. In quadded cables the total coupling loss may be as low as 100 db at 256 kc at which frequency the level difference between the two carrier terminals may be over 60 db.

The longitudinal loss through the office is increased by the presence of repeating coils and voice frequency repeaters on the noncarrier pairs, and also varies with the degree of interconnection between the noncarrier facilities in the two cables. While the metallic circuit and phantom circuit losses may be in the order of 50 to 100 db, the ground return loss will generally be in the order of 10 to 20 db. If the total loss in the interaction path is inadequate, it may be increased by inserting longitudinal suppression coils in the noncarrier pairs. Care must also be taken to insure adequate physical separation of the carrier terminals and the associated wiring. At present it is believed that this will be achieved by using separate, even though adjacent, bays for the two groups of terminals and by separating the wiring by the width of at least one bay.

CONCLUSION

Most of the *N1* carrier noise problems were anticipated in connection with the original development and the engineering of the specific projects. However, in some instances the solutions have been learned from experience, the best teacher of all. *N1* circuits are now operating under conditions considerably more severe than originally expected and the trend will doubtless continue. As for the effectiveness of the control measures, it may be noted that in a recent survey embracing over 4,000 channels, only 8 per cent of the channels exceeded the noise objective and on half the circuits the noise conditions were at least 8 db better than the objective. In a number of cases, certain channels are being kept out of service until one or another of the measures outlined can be applied. As further noise suppression is installed, many of the working channels will also be benefited. The noise survey includes the effect of dial office noise but probably does not fully reflect atmospheric static noise since the measurements were made at random times throughout the year.

No comparable surveys of the amount of crosstalk are available, but discussion with the operating companies, and measurements on some long systems operating in nonquadded cables, indicate that adequate control of crosstalk has also been achieved.

Reference

1. A New Telephone Carrier System for Medium-Haul Circuits, R. S. Caruthers, H. R. Huntley, W. E. Kahl, Ludwig Pedersen. *Electrical Engineering*, vol. 70, Aug. 1951, pp. 692–97.

◆

Discussion

W. C. Farrell (Chesapeake and Potomac Telephone Company of Baltimore City, Baltimore, Md.): We should like to commend the authors for the excellent job they have done in preparing and presenting their paper. The points included are both thoroughly and clearly covered. The following comments on the noise aspects of their presentation which may be of value to interested parties are observations that the Toll Transmission Engineering Section of the General Engineering Department of The Chesapeake and Potomac Telephone Company of Baltimore City has made through the experience of applying the *N-1* systems to the telephone plant in Maryland.

It is suggested that the pairs selected for use for *N-1* systems be those which have good longitudinal-to-metallic balance. This characteristic can be evaluated only by measuring the pairs at the system terminals and at intermediate repeater points prior to their assignment. While this is a straightforward process on nonloaded pairs in a new cable, it can become both involved and expensive in cables working at full capacity with loaded voice-frequency circuits.

This latter situation is the one most frequently encountered in practice. It is often necessary to take circuits out of service to permit unloading and rearranging conductors to make them suitable for carrier use. Obviously, in cases of this type, the selection of pairs by measurement may become quite costly and, unless the testing work is carefully performed, the work operations may have serious effects upon service.

Although our experience to date on interference in *N-1* carrier systems from sources external to the telephone plant has been confined to static and radio transmitters, it is the belief of our organization that the possibility of other carrier systems being originators of future trouble should not be overlooked. As in the cases already outlined, the longitudinal path to ground via taps from cables would afford this source an entrance path to the *N-1* cable. This belief is based on the fact that the frequency band utilized for the *N-1* system is also used by right-of-way companies. For instance, the power companies are using frequencies from 7 kc to 245 kc. It seems advisable to keep in mind this fact in laying out new carrier systems or investigating existing interference problems in the telephone plant.

Administration of suppression methods that have not been standardized, and consequently are not included in formal instructions, has proved to be a problem where large numbers of people from various groups have jobs to perform on the *N-1* cables. This has been the case along the Annapolis-Hyattsville route in our territory. We believe that a large part of this is the result of the limited number of suppression units available during the experiments which necessitated leaving taps untreated that were not important offenders, even though they were in a zone known to include the troublesome exposures. As a result, as these locations are visited and changes are made which may affect the interference level, suppression is seldom added. Also the omission of suppression on some taps tempts the man without a background in inductive interference to be lax in observing informal instructions on the subject. We believe, however, that when the new retardation coil is available in sufficient quantities to meet our needs along with instructions for its use, this problem will be alleviated.

Mention was made of a possible stop-gap measure referred to as a slot rejection filter. To date this has not been used in our territory for the reasons stated in the paper. Within recent weeks, an additional anticipated objection to the slot filter was justified. This is simply the fact that there is nothing to prevent the originator of the radio-frequency interference from shifting his frequency.

For this reason, the interference that was formerly on channel 7 of the system in the Annapolis area now appears on channel 10 at a different voice frequency. This obviously means that the investment in the slot filter has been lost. The suppressors, on the other hand, will be effective at the new frequency.

W. F. Potter (New England Telephone and Telegraph Company, Boston, Mass.): In view of the extensive application of *N* carrier systems, this paper, discussing the control of noise and crosstalk in these systems, is important and most timely. It may be of interest to review briefly some specific experiences encountered in the New England area which illustrate various points covered by the paper.

Secondary induction from telephone plant, particularly noise in the carrier frequency range caused by operation of switching equipment, has been a major problem, being particularly severe where high-low repeaters are located in step-by-step dial offices. On one of our first *N* systems where the *N* carrier base circuits to a high-low repeater were in the same nonquadded cable as the dial exchange lines, a severe noise condition was eliminated by removing the *N* pairs from the exchange cable and placing them in the quadded core of a relieving toll cable for a distance of 700 feet from the office.

In another case, involving an 8-mile repeater section to a high-group receiving terminal, 3,000 feet of compartmented entrance cable was installed in which a quadded core for the *N* carrier pairs is shielded from the dial exchange pairs by layers of steel and copper tape. In another case, where the repeater spacing was slightly over 7 miles, segregation for a distance of 1,300 feet at a high-low repeater point was accomplished by providing a cable with a small quadded core to which the *N* pairs were assigned, using the surrounding nonquadded pairs for dial exchange lines and trunks. Because of the shorter repeater spacing, giving higher carrier levels at the input to the high-low repeaters, no shielding was required around the core.

In laying out *N* systems every effort is made to make dial offices low-high repeater points. Where this arrangement is not economically feasible, repeater spacings adjacent to high-low repeaters have been shortened.

On Cape Cod, two *N-1* systems were installed between Hyannis and Provincetown, a distance of about 50 miles. This cable has about 1,000 drop wires and 400 open-wire leads between these two points. Three channels in each system were adversely affected by the signals from three different radio stations. Suppression of the open-wire and drop-wire leads was not felt to be practicable, and other mitigative measures were sought.

One of these radio stations, approximately 30 miles from Hyannis, and operating on 194 kc with 1,000-cycle tone telegraph signals, interfered with channel 4 of each system. A steady tone of 2,000 cycles was received on channel 4 as a result of beating of the 194-kc carrier of the radio station with the channel frequency of 192 kc. In addition, the telegraph signal at 1,000 cycles was also present on the channel but at a much lower level. By inserting a 2,000-cycle slot filter in the voice-frequency path from the carrier channel to the switchboard at each end of the circuit, the following results were obtained.

On one system at Hyannis the over-all noise without suppression was 55 decibels adjusted (dba) at the zero level point. With the 2,000-cycle slot filter this was reduced to 34 dba. The 1,000-cycle tone was reduced from 39 to 33 dba. The 2,000-cycle noise without suppression was 54 dba. With the 2,000-cycle slot filter this was reduced to a negligible value which could not be measured. Similar results were obtained on the second system at Hyannis and on both systems at the Provincetown end.

A second radio station about halfway between the ends of the carrier systems operating on 406 kc, with continuous-wave code transmission, interfered with channel 5 of each system. Since this frequency lies outside the channel frequencies, a different approach was used to mitigate the noise conditions. The 406-kc radio carrier beating with the 304-kc oscillator in the high-low repeaters produced telegraph code signals at 2 kc. Trial insertion of a 2-kc slot filter at the Hyannis switchboard reduced the 2,000-cycle code signal from 77 dba to 48 dba. This value of 48 dba was not of course satisfactory, so a radio frequency filter tuned to 406 kc consisting of two antiresonant series elements and a shunt seriesresonant element with the midpoint grounded, was tried at a high-low repeater.

This was not effective until the filter was isolated from the line by longitudinal chokes at the input and carrier frequency repeating coils at the input and output of the filter. The simplexed d-c path supplying power to the repeater was by-passed around this filter through a retard coil.

This filter arrangement was used at the two high-low repeater points nearest the radio station, and the following results were obtained at the switchboards.

At Hyannis, without the filter, the noise was 75 dba on both systems. This filter reduced the noise to 25 dba on one system and to 12 dba on the other. At Provincetown, noise of 64 dba on both systems was reduced to 31 dba on one system and to 32 dba on the other.

The third radio station transmitting on 137 kc produced a telegraph code signal of 1,000 cycles on channel 1 of each system, the frequency of channel 1 being 136 kc. No suppression of this interference was attempted with slot filters because of the high level of the code interference with only 30-decibel suppression expected from the slot filter.

The interference to channel 1 and channel 5 has now been eliminated by a change in the transmitting frequency of the two radio stations. The 137-kc transmitter was changed to 147.5 kc which is 7.5 kc above the channel 1 frequency band. Channel 5 is no longer affected by radio interference, since the 406-kc radio station has been changed to 436 kc, and the beat frequency with the 304-kc oscillator now lies between channel 1 and channel 2.

The recently developed optional channel 13 operating at 40 kc and 264 kc will be substituted for channel 4 on all systems in this area so that we will be able to operate all 12 channels of all systems without radio interference.

Crosstalk has not been a problem on any *N-1* systems in the New England company's area.

One problem yet to be solved is that of impulse noise of a few milliseconds' duration, which, although of such short duration and low amplitude as not to interfere with speech transmission, is of such a character as to interfere with carrier telegraph systems and some subscriber special service lines.

A-C Null-Type Recorder with Balancing Amplifier Which Provides Damping and Suppresses the Quadrature Component

A. J. WILLIAMS, JR.
MEMBER AIEE

J. F. PAYNE, JR.
ASSOCIATE MEMBER AIEE

ELECTROLYTIC conductivity is a good example of a measurement which can be made most satisfactorily by using an a-c null-type instrument. The use of alternating current instead of direct current reduces polarization at the electrodes. The use of a null-type recorder permits the measurement of conductivity directly by means of a Wheatstone bridge, thus avoiding the need for voltage regulation. This paper describes difficulties encountered when an unimproved electronic balancing amplifier is tried for this type of instrument. Theory and performance of an improved amplifier are given.

Electrolytic conductivity measurements are widely used in industry to determine chemical concentration. This measurement of concentration may be used to control the product strength (as in the production of sulphuric acid) or to determine the leakage or carry over of a solution of one concentration into a solution of different concentration (as in the contamination of condensed steam from a steam turbine).

Fig. 1 shows a Wheatstone bridge arranged for the measurement of electrolytic conductivity. If the filled electrolytic conductivity cell behaves as a pure resistance R, the bridge may be completely balanced by adjustment of the opposite arm R_1. The bridge equations show that

$$1/R = R_1\frac{1}{R_2R_3}$$

hence the conductance of the cell $1/R$ is directly proportional to R_1.

Actually a filled electrolytic conductivity cell may not behave as a pure resistance, and its equivalent circuit may include a series capacitive reactance, as shown in Fig. 2. In this case, adjustment of R_1 cannot completely balance the bridge and the phasor diagram of Fig. 2 is an attempt to explain this pictorially. The voltage remaining across the detector E_{db} has a phase in quadrature with the desired component of detector voltage. The desired component is that component of detector voltage produced by a change in electrolytic conductivity.

This quadrature component does not interfere with the balancing of a-c null-type recorders of the form which use an a-c galvanometer and a mechanical balancing system. For this form of recorder the field of the a-c galvanometer is phased so that the quadrature voltage produces a torque on the moving coil which averages to zero.[1,2] Otherwise stated, the effect of the quadrature component is suppressed.

In 1933, a d-c null-type recorder was produced commercially, in which a reversing motor and an electronic balancing amplifier was used.[3,4] Through the years the electronic balancing system has been simplified[5] and refined[6] but when modified for a-c measurement its performance has been inferior to that of the a-c galvanometer and the mechanical balancing system, when used where a quadrature component of bridge output exists, as in the measurement of electrolytic conductivity. An electronic balancing amplifier which is improved to be free of this inferiority is the basis for this paper.

The idea for the improved amplifier evolved from a study of difficulties with the unimproved electronic balancing amplifier, and from a comparison with the galvanometer form of balancing system. The latter system suppresses the quadrature component prior to amplification, whereas the unimproved electronic system amplifies the quadrature component and relies upon its suppression by the characteristics of the 2-phase balancing motor at the amplifier output. This motor has no rotating field and hence no starting torque, if the voltage applied to the amplifier winding is in time phase with the voltage applied to the other winding normally energized from the line. In practice, two difficulties are encountered:

1. Because of the change in characteristics of the several tubes and other elements in the amplifier with time, temperature, signal

---

Paper 53-244, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 17, 1953; made available for printing April 16, 1953.

A. J. WILLIAMS, JR., and J. F. PAYNE, JR., are with the Leeds and Northrup Company, Philadelphia, Pa.

The dual vibrator was developed by R. E. Tarpley and W. H. Packer. Experimental work during the early stages of the amplifier development was done by W. R. Clark and the late F. F. Dern, Jr.

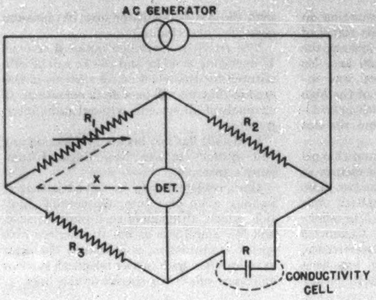


Fig. 1. Wheatstone bridge for the measurement of electrolytic conductivity

If the bridge is completely balanced by adjustment of R_1, its value is then a measure of the conductance $1/R$, since R_2 and R_3 are held constant

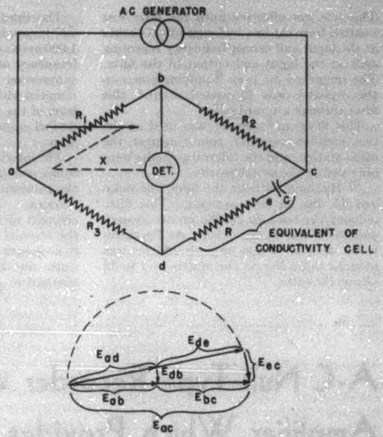
$$\frac{R_3}{R_1}=\frac{R}{R_2}$$

$$R_2R_3=RR_1$$

$$1/R=R_1\frac{1}{R_2R_3}$$

Fig. 2. Source of quadrature component

The equivalent circuit of an actual filled electrolytic conductivity cell may include a series capacitive reactance; hence adjustment of R_1 cannot completely balance the bridge. The voltage remaining across the detector has a phase in quadrature with the component which would be produced by a change in electrolytic conductivity



amplitude, and line voltage, the quadrature component generally applies a voltage to the motor which is not in phase with the voltage applied from the line. A torque results which can only be balanced by a shift of the slidewire and pen away from the correct position. The recorder reading is thus in error.

2. Because it is amplified, the quadrature component causes useless heating in the motor, and loads the later stages of the amplifier so that the desired signal finds an amplifier with reduced power-handling capacity and with modified gain. The result is generally a slower and less sensitive recorder.

The idea for improving the balancing amplifier was basically to suppress the quadrature component prior to amplification.

Systems analogous to the a-c galvanometer were considered. In the a-c galvanometer, the quadrature component produces a torque of double frequency but its average value is zero.[2] The mechanical filtering action of the system suppresses the effect of the double-frequency torque. Analogous systems may be nonmechanical, but they still require a filter (which may also be nonmechanical) to suppress the double-frequency effect resulting from the quadrature component. Filtering introduces delay, which is undesirable in a balancing amplifier for a fast recorder, hence all systems analogous to the a-c galvanometer were classified as undesirable.

The virtue of the a-c galvanometer seems to reside in the single-valued nature of its output. To be more exact, it is the single-valued nature of the output from the periodic pointer-sensing mechanism[7,8] on the galvanometer. Even if there is a quadrature component and even if the magnetic field is not phased to bring the average resulting torque to zero, there is still only one value of output. Therefore the mechanical balancing system has only one input which, by balancing the slidewire, it can reduce to zero or nearly to zero.

In the improved electronic balancing amplifier, the input to the first stage tube is reduced to a single valued function by sampling the unbalance at preferred times on each cycle. The preferred times are at the maximum and minimum of the desired component, and hence at the zeros of the quadrature component. When successive samples are equal, no signal is fed into the first stage tube. When they are unequal, an a-c voltage is fed into the first stage, the polarity of this voltage indicating the sense of the unbalance.

Fig. 3 shows the arrangement of circuit elements used to accomplish this result. The bridge to be balanced is at the left, and at the right is the first stage of the amplifier whose output feeds one winding of a 2-phase motor (not shown) which can drive the slidewire R_1 toward the balance point. The moving contact on R_1 carries along with it a pen for recording (not shown).

The bridge output is sampled by disconnecting capacitor C_{t1} from the bridge by the break of connection between con-

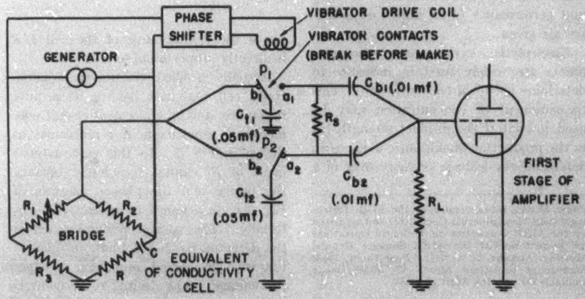


Fig. 3. Simplified circuit

C_{t1} and C_{t2} sample the bridge output at the instant of contact break, thus giving a single valued input to first stage of amplifier. The phase shifter is set to suppress quadrature component. C_{b1}, C_{b2}, R_L and R_B provide damping

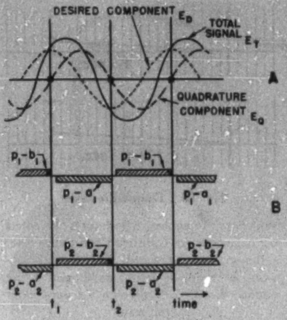


Fig. 4. Timing of contacts

The upper part of the figure shows the total signal and the two components into which it may be resolved

The critical contact times are at t_1 when p_1-b_1 opens, and at t_2 when p_2-b_2 opens. t_1 and t_2 are at the zeros of the quadrature component

tacts p_1 and b_1. These contacts are parts of a switch driven in synchronism with the bridge supply, marked generator in Fig. 3. The phase shifter is set so that the break occurs when the quadrature component of voltage is passing through zero. Shortly after this break, connection is made between contact p_1 and contact a_1. The sample of voltage, retained by capacitor C_{t1}, is impressed upon the amplifier input through capacitor C_{b1}. After the completion of a half cycle, the bridge output is again sampled, this time by disconnecting capacitor C_{t2} by a break between the contacts p_2 and b_2. These contacts had made connection at or about the same time that p_1 made connection with a_1. Since just 1/2 cycle was allowed to elapse, the second sample is taken when the quadrature component of voltage is again passing through zero.

This sequence of events is presented pictorially in Fig. 4. Time is the abscissa, with later times toward the right. The shaded areas at the bottom of the figure indicate contact connections as labeled. The critical times (emphasized with short heavy lines) are when p_1—b_1 opens at time t_1 and when p_2—b_2 opens at time t_2. Preferably, these openings should occur as the quadrature component passes through zero (emphasized by small circles), and Fig. 4 is so drawn. (The total signal will generally contain both components, the desired component and the quadrature component). This sampling technique passes along a signal in which the desired component is enhanced (since the sampling is taken at its peaks) and in which the quadrature component is suppressed (since the sample is taken at its zeros).

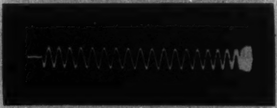


Fig. 5. Input signal to balancing amplifier

This is a cathode-ray oscillogram of a 60-cycle-per-second signal, manually initiated about 4 cycles after manually triggering the single sweep

Demonstration of Quadrature Suppression

The suppression of the quadrature component can also be presented pictorially by cathode-ray oscillograms made with a balancing amplifier including the circuit of Fig. 3. For experimental purposes, the bridge of Fig. 3 was replaced by a simple resistive voltage divider and a manually operated key was added, so the signal to the amplifier could be abruptly initiated. A cathode-ray oscillograph, arranged with d-c amplification, a manually triggered sweep, and a camera, was connected across the input to the balancing amplifier. The two manual operations were coordinated into a single manually operated tap key, so that the sweep was started when the moving arm left its top position and the signal was initiated when the arm arrived at its bottom position. Fig. 5 is the oscillogram. There is no significant signal for a time corresponding to about 4 cycles; then the 60-cycle signal starts abruptly and continues at constant amplitude. The wave length is constant until the spot slows down near to the right hand limit of its travel. This is the type of signal used for the input to the balancing amplifier for the several oscillograms which follow.

Fig. 6 is intended to show the effectiveness of quadrature suppression. To the improved balancing amplifier, including the sampling circuit of Fig. 3, there was applied an input signal of the form shown in Fig. 5. The four oscillograms were all taken at the grid of the output tube of the amplifier. The input signal for all four oscillograms had the same amplitude, but its phase was 90, 30, 10, and 3 degrees, from quadrature, reading from the top trace down. The successive reduction in amplitude from the top oscillogram down shows that the signal amplitude (at this point in the amplifier) approaches zero as the phase of the amplifier input approaches true quadrature. At true quadrature, the zeros of the input signal voltage occur just as the samples are taken, as shown in Fig. 4.

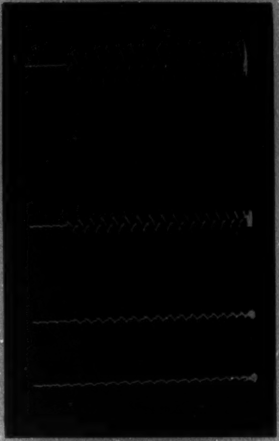


Fig. 6. Quadrature suppression

The input to the amplifier had the same amplitude for all four oscillograms. Its phase was 90, 30, 10, and 3 degrees, reading from the top trace down. The oscillograms (taken at the grid of the output tube of the amplifier) show that the signal amplitude approaches zero as the phase of the input approaches quadrature

Damping

The effectiveness of quadrature suppression is independent of the values of the circuit elements C_{t1}, C_{t2}, C_{b1}, C_{b2}, R_g, and R_L, shown in Fig. 3. For simplicity, the oscillograms of Fig. 6 were made with R_g at zero ohms. This parallels C_{b1} and C_{b2}, so a single capacitor could have been used. The oscillograms would have been essentially the same if C_{b1} and C_{b2} had been short circuited and R_L had been open circuited. In brief, the circuit elements between the synchronous switch and the first-stage tube are not needed for quadrature suppression.

Fig. 7 shows the response of the recorder to an abrupt change of input with the balancing amplifier improved only for quadrature suppression. The top record made on a fast moving chart shows overshoot and oscillation. Although overshoot is tolerable on many feedback-control systems, it is intolerable on recorders because it may erroneously be interpreted as overshoot of the measured quantity, especially with a slow chart as shown at the bottom of Fig. 7. Alarms might erroneously be triggered regardless of

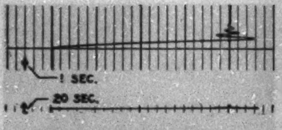


Fig. 7. Underdamped recorder

A fast chart was used for the top record to show overshoot and oscillation. With a slow chart, as in the bottom record, the overshoot of the recorder might erroneously be interpreted as overshoot of the measured quantity, $R_S=0$, $R_L=1$ megohm. Each division is 1 per cent of full scale

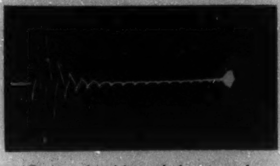


Fig. 8. Amplifier with damping only

In response to a modulated step input, the balancing amplifier gives an output which rises abruptly and then decays to zero, when R_S is infinite

chart speed, because of overshoot. For Fig. 7 the circuit values (not already given in the circuit of Fig. 3) were $R_S=0$ and $R_L=1$ megohm. These were the values used for the oscillograms of Fig. 6.

A study of Fig. 6 by someone skilled in the art of feedback-control systems would lead to the prediction of overshoot, since there is nothing in the amplifier characteristic to offset the adverse effect of inertia in the balancing motor.

The circuit of Fig. 3 was devised to give to the balancing amplifier a characteristic which provides damping for the recorder. With R_S infinite (and R_L still 1 megohm) the response of the amplifier at the grid of the power tube is shown in Fig. 8. The input to the amplifier was a modulated step input, as already shown on Fig. 5. The response has an abrupt rise and then decays rapidly toward zero, the time constant (of the envelope) being about $1\frac{1}{2}$ cycles or 0.025 second.

In evaluating this response, the ideal response for pure "rate action" should be considered. Since the change in input is a step, its rate initially is infinite but of infinitesimal duration; in brief, it is an impulse of finite value. A measure of the amount of the actual response may be taken as its area; for Fig. 8 this is finite, as it should be. A measure of the quality of the actual response may be taken as its compactness in time, or more exactly the reciprocal of its time constant when the form (of the envelope) is that of Fig. 8.

The form of the response may be explained by referring to Fig. 3. When C_{t1} first connects to C_{b1}, the voltage of C_{t1} is impressed upon the first-stage grid. Immediately, current flows through R_L and C_{t1} developing a voltage across C_{b1} and thus reducing the voltage on the grid. The same action takes place between C_{t2}, C_{b2}, and R_L. The action is cumulative, since there is no way for the charge to leak off of C_{b1} or C_{b2}. As the voltage of C_{b1} approaches that transferred by C_{t1}, the signal to the first grid diminishes toward zero. The time constant of C_{b1} and R_L is 0.01 second, but since charging takes place slightly less than half the time, the effective time constant may be slightly over 0.02 second. This is close to the observed time constant of 0.025 second.

An amplifier with the response of Fig. 8 would have no sensitivity to bridge unbalance and hence would not be generally useful. Fig. 9 shows the amplifier response when $R_S=33$ megohms and $R_L=1$ megohm. The final or steady-state response is not zero but about 20 per cent of the maximum value; hence, the amplifier is useful for balancing a null-type recorder. The damping characteristic is retained, as indicated by the large amplitude in the early part of the response.

The steady-state amplitude is explained by the discharge path through R_S of the capacitors C_{b1} and C_{b2} in series. Even in the steady state, they cannot be charged to the voltage on C_{t1} (and C_{t2}), and hence a voltage equal to the difference is transmitted to the grid of the first stage.

A balancing amplifier with the response of Fig. 9 gives a recorder response as shown at the top of Fig. 10. Overshoot is prevented and there is no oscillation. Balance is not completed in minimum time, since faster balancing is shown at the bottom of Fig. 10, for which R_S was 22 megohms. The amplifier response for this condition was taken but is not shown, since it is very similar to Fig. 9, the steady state amplitude being about 30 per cent of the peak amplitude.

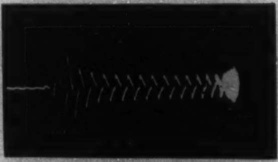


Fig. 9. Amplifier with damping and with sensitivity

In response to a modulated step input, the balancing amplifier gives an output which rises abruptly and then decays to about 20 per cent of the maximum value, when $R_S=33$ megohms and $R_L=1$ megohm

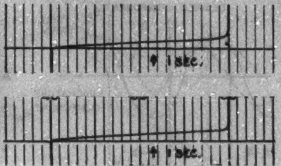


Fig. 10. Damped recorder

With $R_S=33$ megohms, overshoot is prevented but balance is not completed in minimum time, as shown in the top record. With $R_S=22$ megohms, balancing is faster, as shown in the bottom record. Each division is 1 per cent of full scale

When R_L is increased to 4 megohms and R_S is put back at 33 megohms, the steady-state amplitude is about 35 per cent of the peak amplitude and the quality of the damping has been reduced as expected. The recorder with this amplifier does not overshoot but there is a greater tendency toward oscillation, apparently because of the reduced quality of the damping.

GENERAL

The appearance of the recorder is very similar to that of a recorder shown in a previous paper[5]; hence, there is no need for a picture. Fig. 11 shows the balancing amplifier, the appearance of which is essentially the same as that of the balancing amplifier used for the d-c null-type recorder.[6] The rectangular plate at the left covers the hole used by the capacitor for the input filter of the d-c amplifier.[6] The dual vibrator which makes possible the improved performance is shown with its cylindrical case in Fig. 11, and without its case in Fig. 12. It is designed so that the extra mechanical load imposed by the extra set of contacts does not result in poor timing.

Besides electrolytic conductivity, there are other measurements which can be made advantageously with an a-c null-type recorder with the improved balancing amplifier. Power-line frequency measurement by means of a Wheatstone bridge using fixed capacitors, fixed resistors, and variable resistors[9] is one example. Another example is the measurement of the position of a magnetic core in relation to a coil. This device, sometimes referred to as a differential transformer, has found use in the measurement of level, pressure,[10] flow, etc.

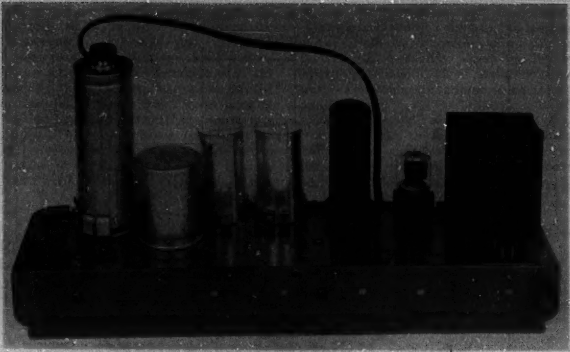


Fig. 11. Balancing amplifier

Appearance resembles that of the more commonly used balancing amplifier for d-c null-type recorder. The vibrator (converter) used for sampling receives its driving power through the cord connection at its top

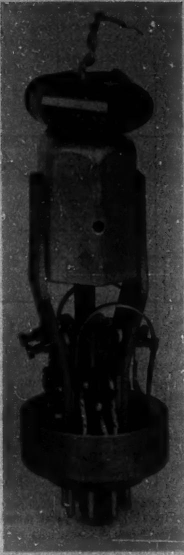


Fig. 12. Dual vibrator

The case has been removed to show the arrangement of contacts and the four adjusting screws. Only three of the six connecting wires show in this view

Although the simple theory for these circuits indicates that the voltage across the detector can be reduced to zero at balance, actual experience shows that there is generally a residual voltage which cannot be eliminated by adjustment of the circuit elements under the control of the balancing system. One component of this residual voltage is the quadrature component, already mentioned in connection with the electrolytic conductivity recorder. A second such component is direct current, but this feeds no signal to the first-stage tube, since successive samples are equal. For the same reason all even harmonics are suppressed regardless of phase. Odd harmonics are suppressed if they have zeros at the sampling times. If odd harmonics are present and their zeros are not at the sampling times, a signal is sent which runs the balancing motor until successive samples are equal, and the input to the first stage is again reduced essentially to zero.

Summary

An attempt has been made to describe the need for and the means used to achieve improvements in the electronic balancing amplifier for the a-c null-type recorder, as applied to difficult measuring problems such as electrolytic conductivity.

By the use of a sampling technique, involving a vibrator type of synchronous switch, the quadrature component of bridge unbalance is suppressed along with other undesired components. The sampling technique is effective because of the single valued nature of the derived output.

Recorder damping is accomplished by an unconventional use of resistive and capacitive elements in conjunction with the switching action of the vibrator.

Response curves, experimentally obtained, are given for the balancing amplifier alone and for the complete recorder. The curves are an attempt to explain pictorially the damping action, as well as quadrature suppression.

References

1. Measurement of Electrolytic Conductivity, **G. F. Rucker.** *Instruments*, Pittsburgh, Pa., vol. 25, Oct. 1952, pp. 1418–19, 1462–63.

2. A Study of Electromagnet Moving Coil Galvanometer for Use in Alternating Current Measurements, **Ernest Weibel.** *Paper No. 297, Bulletin*, National Bureau of Standards, Washington, D. C., vol. 14, 1918–19, pp. 232–58.

3. New Instruments (Speedomax). *Instruments*, Pittsburgh, Pa., vol. 6, Nov. 1933, p. 211.

4. Combined Thyratron and Tachometer Speed Control of Small Motors, **A. J. Williams, Jr.** *AIEE Transactions (Electrical Engineering)*, vol. 57, Oct. 1938, pp. 565–68.

5. Electronically Balanced Recorder for Flight Testing and Spectroscopy, **A. J. Williams, Jr., W. R. Clark, R. E. Tarpley.** *AIEE Transactions (Electrical Engineering)*, vol. 65, April 1946, pp. 205–08.

6. Electronic Recorder with Range and Precision Adequate for the Platinum Resistance Thermometer, **Albert J. Williams, Jr.** *AIEE Transactions*, vol. 71, pt. I, Sept. 1952, pp. 289–95.

7. Electrical Recorder, **M. E. Leeds.** United States Patent No. *1,125,699*, Jan. 1915.

8. Electrical Measuring Apparatus, **L. Y. Squibb.** United States Patent No. *1,935,732*, June 1932.

9. System of Frequency or Speed Measurement and Control, **F. Wunsch.** United States Patent No. *1,751,538*, March 1930.

10. Electric Telemetering System, **J. F. Quereau, A. J. Williams, Jr.** United States Patent No. *2,050,629*, Aug. 1936.

◆

Discussion

Thomas J. Higgins (University of Wisconsin, Madison, Wis.): This well-written and well-integrated paper comprises an excellent account of the difficulties besetting solution of the particular problem in a-c measurements outlined therein and of the means employed for solving it. The use of appropriate sampling to eliminate, in the large, the undesired quadrature component is an ingenious approach to obviation of this "intruding" component. In the discusser's opinion, this paper encompasses a very useful and widely applicable contribution to the domain of a-c measurement techniques.

William G. Amey (Leeds and Northrup Company, Philadelphia, Pa.): Phase discrimination usually is obtained by means of a modulation process which converts the input signal of fundamental frequency into an output signal containing two or more frequency components. If the modulation process is suitable, the amplitude of at least one of those components is determined by

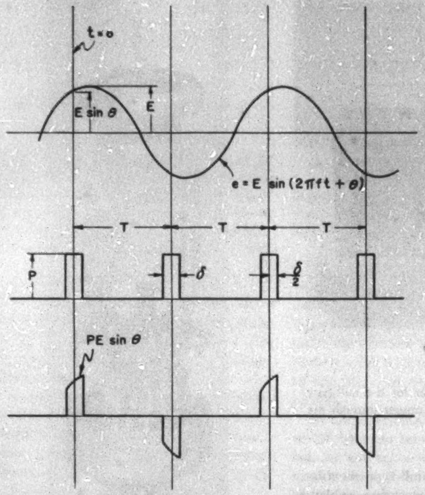


Fig. 13. **Simple pulse sampling for phase discrimination**

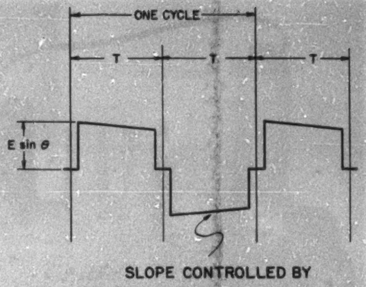


Fig. 14. **Amplifier grid voltage resulting from the use of sampling capacitor**

Note: Wave shape is independent of signal sampled

the relative phase of the input and modulating signals, and thereby provides phase discrimination. For that purpose, the authors have used an interesting scheme of pulse sampling which avoids one of the shortcomings of most pulse-modulation techniques.

Fig. 13(A) shows a generalized signal component having a phase angle θ, such that the instantaneous magnitude is $E \sin\theta$ at time $t=0$. For purposes of phase discrimination, this component can be modulated by sampling pulses of the type shown in Fig. 13(B) to give a resultant modulated output signal like that of Fig. 13(C).

If the sampling time δ is made sufficiently small compared to the fixed repetition time T (that is, δ/T having a value of 0.01 or less), the resultant output pulses can be considered to be flat-topped with a pulse height of $PE\sin\theta$. The instantaneous values S of this wave form are given quite accurately by the Fourier series

$$S \cong 2EP\frac{\delta}{T}\sin\theta\sum_{k=1}^{\infty}\cos 2\pi(2k-1)ft \qquad (1)$$

where f is the fundamental signal frequency $=1/2T$, and t is the time measured from the time reference.

According to equation 1, this output signal consists of a component of fundamental frequency f, and other components which are odd harmonics of f. All components have the same amplitude, but only the fundamental is useful for the authors' purposes. As expected, the amplitude varies with the relative phase of the sampling pulse and the input signal component. An in-phase component corresponds to $\theta=90$ degrees and gives the maximum signal output, as shown in Fig. 6. As the quadrature condition is approached, θ decreases; at true quadrature, the output is zero and phase discrimination is complete.

Such phase discrimination is purchased, however, at an intolerable price. The quantity E represents the smallest input signal amplitude with which the authors would like to deal. P is the sampling-pulse amplitude, which has a fixed value of unity when the sampling is done by switching, as in this case. Consequently, as δ is reduced to provide better phase discrimination, the useful output signal amplitude decreases proportionately. Some compensation is afforded by the use of an amplifier of higher gain and lower noise level than normal. In general, however, this simple method of sampling leads to excellent phase discrimination, with virtually no output signal to prove it.

Recognizing this difficulty, the authors used a more suitable approach to the problem. Instead of merely connecting the amplifier briefly to the bridge output, they used the energy stored in a sampling capacitor. The capacitor was connected to the bridge until the desired instant for sampling was reached. The capacitance was so related to the bridge impedance that the capacitor voltage followed that of the in-phase component very closely. With exactly the right phasing, the capacitor was disconnected when it was charged to the peak voltage of the in-phase component. This provided excellent phase discrimination, since the sampling time was effectively only that required to disconnect the capacitor from the bridge.

The capacitor, with its charge proportional to the peak value of the in-phase component, was shortly thereafter connected to a network on the input grid of the amplifier. It was left in this connection as long as practicable before being switched back to the bridge connection. A duplicate capacitor and switch were so connected and phased that effective sampling of the other half cycle of the in-phase component was also accomplished. With a suitably high impedance of the grid network, the voltage on the input grid was then approximately that shown in Fig. 14. Fourier analysis of this waveform reveals that the desirable fundamental component has essentially the same magnitude as the original in-phase signal component which was sampled. The reason is that the energy stored in the intermediate sampling capacitor is made available to spread the short-duration pulses of Fig. 13(C) into those of longer duration in Fig. 14.

This desirable result is obtained at the cost of somewhat greater loading of the bridge by the sampling capacitor than would have been true in the simple sampling arrangement first discussed. There appears to be no adequate means of compensating the difficulties of the simple scheme, however, while the effects of the authors' sampling capacitor can be compensated by suitable bridge design or by means of impedance-matching transformers.

George B. Hoadley (North Carolina State College, Raleigh, N. C.): Dr. Amey has emphasized how the authors' use of the two capacitors C_{t1} and C_{t2} has allowed them to use a sampling technique without losing amplitude of the fundamental signal input to the amplifier. I wish to point out that the price paid for this is a reduced input impedance to the amplifier, which comes about because the bridge must supply charging current to the capacitors. This lowered input impedance is not objectionable at all in the use of the amplifier with the conductivity bridge, and the authors are to be congratulated for their ingenuity in developing this system.

A. J. Williams, Jr., and J. F. Payne, Jr.: The authors appreciate the discussion of Professor Higgins. The word "intruding," which he uses with reference to the quadrature component, expresses very well our sentiments concerning this undesired component.

Dr. Amey shows mathematically how the amplitude of the fundamental is reduced by the sampling technique when sampling is accomplished by simply connecting the first stage of the amplifier to the bridge for brief periods during each cycle. Sampling by the use of recurrent but brief periods of connection through a metallic switch was used by Rosa[1] in an instrument for the analysis of recurrent wave forms prior to the development of the cathode-ray oscillograph. This approach was considered for the recorder problem, but was discarded because of the complexity of the vibrator which would be required to provide damping as well as quadrature suppression.

Professor Hoadley points to the loading of the bridge by connection of capacitors C_{11} and C_{12} alternately across the bridge output terminals. This is unquestionably true. Two comments appear to be appropriate.

Because the input impedance to the electron tube is high, the value of the capacitors can be small and the input impedance to the amplifier need not be low. It is about 50,000 ohms for the values shown in Fig. 3.

If one or more of the amplifier tubes were moved to a position ahead of the vibrator, a higher impedance would be presented to the bridge. This arrangement would introduce phase shift and the possibility of phase-shift variations which might not be tolerated.

REFERENCE

1. Edward B. Rosa. *Physical Review*, Cornel University, Ithaca, N. Y., vol. 6, 1898, pp. 17-42

Basic Concepts in the Analysis of Stationary Electric Circuits

D. W. SPENCE
ASSOCIATE MEMBER AIEE

C. R. CAHN
ASSOCIATE MEMBER AIEE

THE NEED FOR and part of the purpose of this paper is demonstrated by the following example: Fig. 1 shows a conducting ring linked by magnetic flux ϕ. Three identical voltmeters are connected symmetrically around the ring as shown. What will be the readings of the voltmeters in terms of the rate of change of the flux $d\phi/dt$ and the ring resistance R?

It is the experience of the authors that many capable and experienced engineers will analyze this device by constructing the incorrect equivalent circuit of Fig. 2. From it they will conclude that the voltages between the nodes are zero and, consequently, that the voltmeters indicate zero. This error is usually due to the engineers not having given sufficient thought to:

1. What—precisely—do they mean by voltage?
2. What—precisely—does a voltmeter indicate about a circuit?
3. How—precisely—does such a circuit work?

Another contributing factor is that the usual electrical engineering fundamentals course, in connection with stationary circuit analysis, is nothing but electrostatics and magnetostatics with Faraday's law tacked on. The analysis of all electromagnetic apparatus operating at any frequency above d-c requires, however, electrodynamic fundamentals, including some field concepts, but avoids dependence on electromagnetic field theory.

Many engineers can, of course, correctly solve the problem of Fig. 1. In most cases this is done by superimposing a supplementary procedure onto their commonly used network analysis. The authors believe, on the other hand, that a circuit theory based on the fundamentals presented in this paper leads directly to the routine solution of all stationary circuits.

The specific objectives of this paper are as follows:

1. To review the circumstances under which a potential function may be defined and the usefulness thereof;
2. To provide definitions of potential difference, electromotance (emf), and what we call the Ri (resistance-current) product, which are of sufficient generality and precision to be useful in near-zone circuit analysis at all frequencies and which differentiate between these quantities;
3. To emphasize the importance of the $q(-\partial \mathbf{A}/\partial t)$ force in electric circuits at all frequencies at which transformer action is significant;
4. To maintain that Kirchhoff's second law is legitimately expressed in two different forms, and to show how and under what circumstances each form is applied.

The present development is restricted to stationary circuits in the near zone.

Conservative and Nonconservative Force Fields

In the analysis of electric circuits, the distribution of currents, the work done, and the energy interchanged in the process of charge flow are of interest. Hence, we are concerned with forces and line integrals of forces and, since most electric forces are experienced through regions of macroscopic dimensions, with fields of force. Of great importance is whether or not a force field or a given component of a force field is conservative. Mathematically speaking, if $\mathbf{f}$ constitutes the force field under consideration, the field is conservative in a given region if and only if

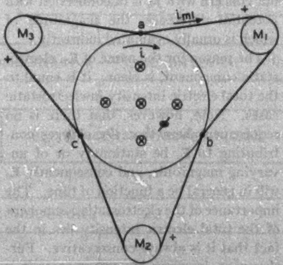


Fig. 1. A flux-linked ring

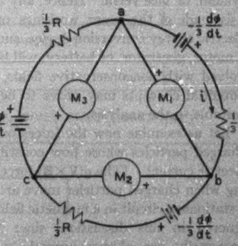

a
$\frac{1}{3}R$
$-\frac{1}{3}\frac{d\phi}{dt}$
+
M3
M1
+
+
i
$-\frac{1}{3}\frac{d\phi}{dt}$
+
$\frac{1}{3}R$
c
M2
+
b
$\frac{1}{3}R$
+
$-\frac{1}{3}\frac{d\phi}{dt}$

Fig. 2. Incorrect equivalent circuit of Fig. 1

Paper 53-268, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted Jan. 19, 1953; made available for printing April 30, 1953.

D. W. SPENCE is with Syracuse University, Syracuse, N. Y., and C. R. CAHN is with Syracuse University and the Niagara Mohawk Power Corporation, Buffalo, N. Y.

The writers wish to record their appreciation to members of the Electrical Engineering Department, Syracuse University, Syracuse, N. Y., and in particular to Dr. W. R. LePage and Dean R. A. Galbraith, for many discussions over the last 3 years concerning the concepts in this paper.

$$\oint \mathbf{f}\cdot d\mathbf{l}=0 \qquad (1)$$

around any and all conceivable closed paths in the region. If, however, the field is not conservative, a conservative component of the field may be defined; that is, equation 1 is satisfied for the component. Physically speaking, a force field is conservative in a region if a body on which the force acts would experience no net interchange of energy with the field if it were to pass around any arbitrary closed path in the region back to the point from which it started, in negligible time.

For a conservative force field, it is possible to define a scalar potential function, as is commonly done for electrostatic fields. Such potential functions are extremely useful because of two properties they possess:

1. There is a unique potential associated with each point in the region of the field;
2. The potential difference between two points is a measure of the energy interchange associated with motion from one point to the other, by any path whatever, of a body on which the force acts.

For a conservative component of a nonconservative field, it again is possible to define a unique potential function. The corresponding potential difference in general is not a measure of total energy interchange in such a case. If, on the other hand, an attempt were made to define potential difference in terms of the total energy interchange, the "potential difference" between two points in a nonconservative force field would not be independent of path, and the "potentials" would not be unique. In the latter two cases the usefulness of the potential function largely vanishes.

The electrical engineer must consider nonconservative force fields, since a current flowing around a closed path in a conservative electric force field, by definition (equation 1), cannot be the agent for the continuous nonreversing coupling of energy from the force field to a nonelectrical (mechanical, chemical, etc.) system, or vice versa. Hence, any analysis made of the inner workings of an electric energy conversion device, such as a motor, generator, or battery, will be involved with nonconservative fields, and potential concepts may prove to be unsuitable to the analysis.

Let us examine now the forces on the charged particles whose flow constitutes electric current. (The $q(\mathbf{V}\times\mathbf{B})$ force acting when charged particles move around a stationary circuit in a magnetic field are ignored in this discussion, since they necessarily are always normal to the filamentary conductors to be considered and hence are of negligible effect in connection with work and energy.) It is convenient to speak not of the force itself, but of the force per unit charge experienced by charged particles in an electric force field, and this ratio with its associated direction is called the electric intensity **E**. Thus

$$\mathbf{E}=\frac{\mathbf{f}}{q} \qquad (2)$$

This definition is made generally for time-varying as well as magnetostatic situations.

The total electric intensity **E** must be resolved into two components, which we will call the electrostatic component $\mathbf{E}_s$ and the electromotive component $\mathbf{E}_m$.

The electrostatic component $\mathbf{E}_s$ is equal to the total electric intensity which would be calculated considering only charges and their positions. The most suitable definition for present purposes is in terms of sources, as follows

$$\mathbf{E}_s=\frac{1}{4\pi\epsilon_0}\left[\int\!\!\int\!\!\int\frac{\rho\mathbf{r}d\tau}{r^3}+\int\!\!\int\frac{\sigma\mathbf{r}ds}{r^3}+\int\frac{\lambda\mathbf{r}dl}{r^3}\right] \qquad (3)$$

where

$\mathbf{r}$ = the vector radius from the element of volume $d\tau$, surface ds, or length dl to the point at which $\mathbf{E}_s$ is being evaluated
ρ = the volume density
σ = the surface density
λ = line density of charge, including all separated or simulated charges in polarized electric media in the region under consideration

The integration is through all volumes, over all surfaces, and along all lines making up the apparatus being analyzed. The extensive information on the evaluation of $\mathbf{E}_s$ in the literature is all applicable, of course. Equation 3 and the literature pertaining to $\mathbf{E}_s$ are mainly of theoretical interest here, however, since in circuits our concern with $\mathbf{E}_s$ is in connection with potential difference, the magnitude of which is usually evaluated indirectly.

The reason for the name of $\mathbf{E}_s$, electrostatic component, is clear; it is equal to the total electric intensity in electrostatic cases. Note, however, that there is no requirement here that the charges contributing to $\mathbf{E}_s$ be stationary or of unvarying magnitude, and consequently $\mathbf{E}_s$ will in general be a function of time. The importance of the electrostatic component of the total electric intensity lies in the fact that it is always conservative. Furthermore, in many important cases it is the total electric intensity in the region of interest even when the field varies with time.

Where the vector difference between **E** and $\mathbf{E}_s$ is not zero, we make the following definition of the electromotive component $\mathbf{E}_m$

$$\mathbf{E}_m=\mathbf{E}-\mathbf{E}_s \qquad (4)$$

so that by definition the two components $\mathbf{E}_s$ and $\mathbf{E}_m$ alone comprise the total electric intensity **E**.

Investigation of physical devices discloses that in practically all regions where an electromotive component exists, it is nonconservative and consequently can furnish energy to or extract energy from an electric circuit carrying a current; hence the name. The forces represented by the electromotive component (that is, the $q\mathbf{E}_m$ forces) are present in all devices which can act as energy sources or as energy sinks (except those which dissipate energy directly as heat); for example rotating machinery, transformers, batteries, thermocouples, and piezoelectric crystals.

In connection with stationary circuits, the nonconservative force or intensity component with which electrical engineers are most concerned is that found in inductors, transformers, and coupled circuits. This intensity component is a fundamental quantity, but it is not encountered in either electrostatics or magnetostatics, where most fundamental forces and intensities are studied. Briefly, whenever the currents in electric apparatus change with time, charged particles in the region of the apparatus experience a force in addition to those they experience under magnetostatic conditions. This force is described at each point by an electric intensity component which we shall call the induced electromotive component. It is a function of the time-rate of change of currents and current densities and is given by an equation analogous to equation 3, as follows

$$\text{Induced }\mathbf{E}_m=-\frac{\mu_0}{4\pi}\left\{\int\!\!\int\!\!\int\frac{\left(\frac{\partial\mathbf{J}}{\partial t}\right)d\tau}{r}+\int\!\!\int\frac{\left(\frac{\partial\mathbf{K}}{\partial t}\right)ds}{r}+\int\frac{\left(\frac{\partial\mathbf{i}}{\partial t}\right)dl}{r}\right\} \qquad (5)$$

where

r = the distance from the element of volume $d\tau$, surface ds, or length dl to the point at which $\mathbf{E}_m$ is being evaluated
$\mathbf{J}$ = the vector current density
$\mathbf{K}$ = the vector surface-current density
$\mathbf{i}$ = the vector filamentary current, including all simulated currents and current

densities in magnetized media in the region under consideration

The integration is through all volumes, over all surfaces, and along all filaments making up the apparatus being analyzed. It follows from equation 5 that the induced electromotive component is simply related to the magnetic vector potential **A** as follows

$$\text{Induced } \mathbf{E}_m = -\frac{\partial \mathbf{A}}{\partial t} \qquad (6)$$

where again the literature on the evaluation of **A** is applicable.

This induced electromotive component of electric intensity usually is not easy to evaluate directly, and if evaluated, may still be difficult to integrate as required for the solution of circuit problems. It consequently is little studied in connection with circuits. For purposes of understanding, it is worth more study of a semiquantitative nature; it indicates, for instance, the amount of the electromotance, or emf, of self-induction developed in any section of a closed loop, and it permits a nice insight into the effect of irregular conductor size around an inductive loop. It is essential to the study of the behavior of conductors which are not part of closed loops. For strictly quantitative purposes, however, the effect of the induced electromotive component is nearly always determined in an indirect manner.

The Essential Circuit Variables

In addition to current, there are three electric variables which are essential to the analysis of electric circuits, and these quantities we define as follows

Potential decrease from *a* to *b*

$$= v_{ab} \equiv \int_a^b \mathbf{E}_s \cdot d\mathbf{l} \qquad (7)$$

Electromotance, or emf, from *a* to *b*

$$= e_{ab} \equiv \int_a^b \mathbf{E}_m \cdot d\mathbf{l} \qquad (8)$$

$$Ri \text{ product from } a \text{ to } b = (Ri)_{ab} \equiv \int_a^b \mathbf{E} \cdot d\mathbf{l} \qquad (9)$$

All of these quantities are measurable in volts. The potential decrease from one point to another is independent of the path of integration. This is not true of the electromotance and the *Ri* product, however, and the paths of integration must be specified in these cases if they are not obvious.

The term "*Ri* product" and the compound symbol $(Ri)_{ab}$ used in the definition of the line integral of the total electric intensity **E** obviously are not as simple and are more specialized in origin than is desirable in such a fundamental definition. They become a little ridiculous in the limiting case of capacitors and open circuits, where *R* approaches infinity and *i* approaches zero with the product continuing to have a finite value not zero. But the introduction of a new term and a new symbol is also somewhat undesirable, and the fact is that in circuit work with filamentary conductors the line integral of **E** is practically always represented by an *Ri* product. **E** itself is proportional to the total force, the vector sum of the forces of the electrostatic and electromotive components, and results in a current flow great enough so that the energy an electron, say, gains from *E* along a particular segment of the path is exactly absorbed by interelectron collisions. This work per unit charge is nearly always proportional to current, and we call the proportionality factor resistance. Better terminology and symbols definitely are needed however.

The avoidance of the terms "increase," "decrease," "rise," and "drop" in the general definitions of the electromotance and the *Ri* product is necessary because such terms have meaning only in connection with potential, and the line integrals of $\mathbf{E}_m$ and **E** cannot be described by potential functions. The very important special case wherein it is meaningful to talk of an "*Ri* drop" will be discussed in the following.

The three definitions, equations 7, 8, and 9, are made more meaningful by the following considerations. In electric conduction processes, three distinct types of energy phenomena are possible. They are:

1. Potential energy may be stored in or withdrawn from the electrostatic component of the electric intensity. The measure of this interchange is the potential difference.
2. Energy may be reversably transferred into or out of the circuit from or to mechanical form, chemical form, magnetic form, thermal form, etc. The measure of this interchange is the electromotance.
3. Energy may be irreversibly dissipated as heat in the conducting medium. The measure of this interchange is the *Ri* product.

Fig. 3 compares the current-producing components of electric force per unit charge which are utilized in this paper. For completeness, the **V×B** term, which can be considered as a component of the total force per unit charge as far as conduction effects are concerned, is included in the table. However, because of the complications involved in evaluating the various components in reference frames other than one which is stationary with respect to all media and apparatus under consideration, the present discussion is limited to stationary cases.

The breakdown of **E** utilized in Fig. 3 is presented on the basis not of uniqueness but of usefulness. The concepts to be used in circuit analysis are developed with the assumption that the charge and current distributions throughout the apparatus being analyzed are known. From this assumption, useful circuit definitions are formulated, and practical methods of evaluating the defined quantities are derived as indicated in the following.

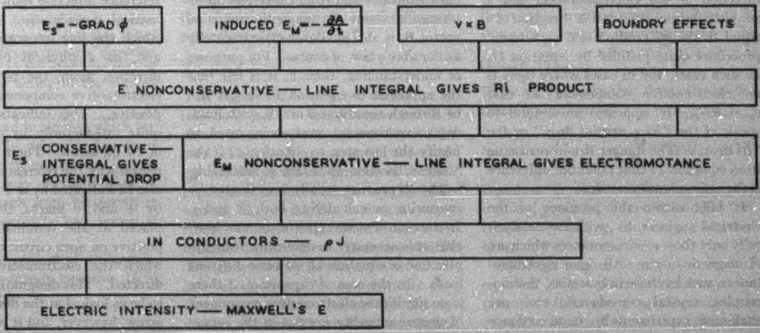

E_S = −GRAD ϕ
INDUCED E_M = −∂A/∂t
V × B
BOUNDRY EFFECTS
E NONCONSERVATIVE — LINE INTEGRAL GIVES Ri PRODUCT
E_S CONSERVATIVE — INTEGRAL GIVES POTENTIAL DROP
E_M NONCONSERVATIVE — LINE INTEGRAL GIVES ELECTROMOTANCE
IN CONDUCTORS — ρJ
ELECTRIC INTENSITY — MAXWELL'S E

Fig. 3. Forces on charged particles

The three quantities defined by equations 7, 8, and 9 usually are not evaluated by means of these defining equations, but by other methods derived from the definitions. Except in capacitors, the potential difference, when it can be evaluated at all, is evaluated by means of well-known equations derived as follows

$$v_{ab} = \int_a^b (\mathbf{E}_s \cdot \mathbf{dl}) = \int_a^b (\mathbf{E} \cdot \mathbf{dl}) - \int_a^b (\mathbf{E}_m \cdot \mathbf{dl}) = (Ri)_{ab} - e_{ab} \qquad (10)$$

More customarily

$$v_{ba} = e_{ab} - (Ri)_{ab} \qquad (11)$$

In the case of capacitors, of course, $v = q/C$. Equations 10 or 11 may be used to evaluate v_{ba} after e_{ab} and $(Ri)_{ab}$ are evaluated, although it may not always be possible to evaluate all the quantities in the equations, as will be demonstrated below.

The fact that the value of potential difference is equal to the difference between the values of the electromotance and the *Ri* product does not, of course, mean that the electromotance and the *Ri* product are themselves potential differences. But there seems to be a powerful, if illogical, tendency to feel that they do mean just that. A somewhat analogous situation arises in mechanics. Consider a body, initially at rest, which is propelled vertically upward from a reference plane in simple translation. Then at any instant the energy relation for the body is

$$\text{potential energy} = \text{work input} - \text{kinetic energy}$$

In spite of this relationship, no one holds that work input and kinetic energy are forms of potential energy, or that it is possible adequately to describe the energy relationships of mechanics by reference to potential energy alone, except in special cases. As a matter of fact, the representation of the special case of a battery by a simple potential rise equal to and in the direction of the electromotance, and a simple potential drop in the direction of *i* equal to the *Ri* product is a very useful procedure quite justified by equation 11. In such cases, and in cases where there is no electromotive component, so that $v_{ab} = (Ri)_{ab}$, it appears meaningful to speak of the "*Ri* potential drop" or the "*Ri* drop." The danger lies in reasoning from equation 11 that potential difference is the only quantity needed.

It falls within the province of the electrical engineer to evaluate analytically only those electromotances which are of magnetic origin. All other electromotances, such as those in batteries, thermocouples, crystal transducers, etc., are evaluated experimentally in accordance with equation 11; that is, with a high-resistance voltmeter. This is possible since they are nearly always practically independent of current within their operating range. The evaluation of electromotances of magnetic origin is very especially the concern of the electrical engineer, however. In such cases, we have from equations 5 and 8

$$e_{ab} = \int_a^b \left(-\frac{\partial \mathbf{A}}{\partial t} \right) \cdot \mathbf{dl} \qquad (12)$$

Although equation 12 is a usable equation, in most cases it is easier to evaluate the time rates of change of the magnetic fluxes linking the circuit loops of which branch *ab* is a component. This information can be utilized analytically as follows.

Although specifically not the definition, for our purposes, of either quantity, the following relation between **B** and **A** can be derived

$$\mathbf{B} = \text{curl } \mathbf{A}$$

Then, with the aid of Stokes's theorem, given a stationary surface S bounded by a stationary closed curve Γ, the electromotance acting around the path Γ is

$$e_\Gamma = \oint_\Gamma \left(-\frac{\partial \mathbf{A}}{\partial t} \right) \cdot \mathbf{dl} = \iint_S \left(-\frac{\partial}{\partial t} \text{curl } \mathbf{A} \right) \cdot \mathbf{ds} = \iint_S \left(-\frac{\partial \mathbf{B}}{\partial t} \right) \cdot \mathbf{ds}$$

$$e_\Gamma = -\frac{d\phi_s}{dt} \qquad (13)$$

where ϕ_S is the total flux of **B** across S, e_Γ acting clockwise as viewed in the reference direction of the flux ϕ_S. (It should be noted here that the method of extension of these concepts to nonstationary circuits is indicated by Cohn's work.[1])

Equation 13, which can be derived from our definitions and from experimental information about the magnetic induction vector **B**, is also available experimentally as Faraday's law, of course. For purposes of understanding, though, it is felt that the approach to equation 13 should still be through equations 5 and 8, with Faraday's experimental work being used to justify the last step to equation 13 if the Stokes's theorem approach is unsuitable.

The *Ri* product usually is evaluated by measuring or calculating both *R* and *i*. In the case of some types of sources, however, it is necessary to determine *R* by application of equation 11 with two different loads. In the case of capacitors, if there is no significant electromotive component of electric intensity present in the capacitor, we have, integrating through the capacitor

$$\int \mathbf{E} \cdot \mathbf{dl} = \int \mathbf{E}_s \cdot \mathbf{dl}$$

$$Ri = v = \frac{q}{c}$$

If the electromotive component is not negligible, but is known, it may still be possible to evaluate *Ri* through use of the complete expression

$$\int \mathbf{E} \cdot \mathbf{dl} = \int \mathbf{E}_s \cdot \mathbf{dl} + \int \mathbf{E}_m \cdot \mathbf{dl}$$

The question might be asked, in connection with this section: What is voltage? It appears in the literature that "voltage" is used to describe all three quantities defined above. In a majority of the cases, it refers to potential difference, except when used in the phrase "induced voltage," when it refers to electromotance. In the future it is suggested that "voltage" be used synonymously with potential difference.

The question might also be asked: What does a voltmeter indicate? A nonelectronic voltmeter, of course, responds to the current through it, and is calibrated to indicate the product of the current and the internal resistance between the voltmeter terminals—in other words, the *Ri* product through itself. An electronic voltmeter responds to and indicates the integral of total **E** between grid and cathode of the input tube, which, again, is defined as the *Ri* product.

The question of the designation of the reference, or positive, direction of the various electric quantities defined above may well be considered in the light of the definitions. The reference direction of the *Ri* product is just the reference direction of the current *i*. The general practice of using double subscripts or else an arrow to indicate this direction seems entirely satisfactory.

In the case of the electromotance, the reference direction indicates, as a mathematical consideration, the direction in which the line integration is performed, and, as a physical consideration, the direction along the path in which the electromotive component is acting when positive. This indication also is given quite satisfactorily by double subscripts or by an arrow. There is no prohibiting reason why the reference direction cannot also be indicated by a + polarity mark or by + and − marks, the + mark being placed at the terminal which becomes positive on open circuit and hence toward which the electromotive component is directed. This designation does not seem to be as logical as the double subscripts or arrow, however, and it probably tends to increase the confusion between electromotance and potential difference.

Polarity marks are very well suited to the indication of the reference direction of a potential difference, on the other hand. A potential decrease from one point to another in a circuit arises only from an excess of charge at the first point, and the principle use of polarity marks is to indicate just such a situation. The use of an arrow to indicate the reference direction of a potential difference is also quite possible if subject to careful definition, but it seems to be little more than an unfortunate carry-over from the electromotance convention. The confusion in the literature arising from the use of arrows to indicate both potential increases and potential decreases seems to emphasize the unsuitability of the arrow for this purpose. As a demonstration of one convention felt to be more satisfactory, the examples in this paper use the single + and brace notation of LePage and Seely.[2,3]

Kirchhoff's Laws

Kirchhoff's first law is not subject to consideration here. Kirchhoff's second law in both forms is a consequence of the conservative nature of the electrostatic component of electric intensity $\mathbf{E}_s$, as expressed in the equation

$$\oint \mathbf{E}_s \cdot d\mathbf{l} = 0 \qquad (14)$$

Consider first the electromotance form of the law. It is derived as follows

$$\oint \mathbf{E} \cdot d\mathbf{l} = \oint \mathbf{E}_s \cdot d\mathbf{l} + \oint \mathbf{E}_m \cdot d\mathbf{l}$$

$$\oint \mathbf{E} \cdot d\mathbf{l} = \oint \mathbf{E}_m \cdot d\mathbf{l}$$

$$\Sigma Ri = \Sigma e \qquad (15)$$

where all integrations and summations are in the same direction around the same closed path. It can be expressed in words as follows:

> "Around any circuit loop the algebraic sum of the *Ri* products, a given direction being taken as the reference or positive direction of current, is equal to the algebraic sum of the electromotances acting in the current reference direction."

This form is required in circuits which are everywhere linked by magnetic flux, as in the analysis of electric machinery. It embodies the terminology originally used by Kirchhoff,[4,5] although from his proofs it appears that he actually had potential concepts in mind.

The second, or potential difference, form of the law is simply equation 14 in circuit terms, that is

$$\Sigma v = 0 \qquad (16)$$

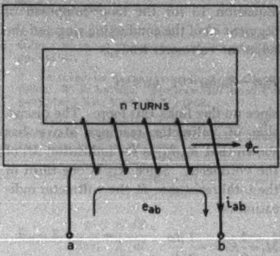


Fig. 4. An inductor with a large number of turns

It can be expressed in words as follows:

> "Around any circuit loop, the algebraic sum of the potential differences is zero."

This form is applicable in circuits composed of lumped circuit elements, including coils and transformers, but with no distributed fluxes linking the circuit loops. It is essential to modern network analysis.

In those many cases where the electromotance in each branch can be evaluated and represented as a simple potential difference, equation 16 is usually simpler and more error-proof to apply. In cases of magnetically induced electromotance, however, it may or may not be possible to evaluate the electromotance in each branch, and so may or may not be possible to evaluate the potential difference between junctions, and to use equation 16. In such cases equation 15 may be required.

Examples

The application of the concepts presented in this paper to the solution of electric circuits is demonstrated in the following three examples. They are also intended to illustrate the mechanics of the circuit action, the energy phenomena, and the obtaining of the potential-difference equivalent circuit.

EXAMPLE 1

In Fig. 4 is an *n* turn coil wound on an iron core (with *n* large), connected in a circuit whose elements are not close to the core. The changing current in the winding and the simulated currents in the core produce an induced electromotive component, $\mathbf{E}_m = -\partial \mathbf{A}/\partial t$, tending to cause charges to move from one terminal of the inductor to the other. Thus, a potential difference is set up between the terminals, and an electrostatic component is produced both inside and outside the inductor. Within the winding, the electrostatic component of intensity opposes the electromotive component, the resultant of the two being just sufficient to maintain the current through the conducting medium. Because of the large number of turns on the inductor, acting to build up the charge separation, the electromotive component of intensity is negligible in comparison with the electrostatic component outside of the immediate vicinity of the core. Therefore, as far as the external circuit is concerned, the potential difference between the terminals is the agency by which external currents are made to flow and by which energy is coupled in or out of the inductor, and it is the measure of energy per unit charge passing the terminals of the inductor. The electromotive component inside the inductor serves the necessary purpose of maintaining the potential difference between the terminals, but it does not directly couple energy to the outside circuit.

The analysis of the inductor uses the concepts in equations 7, 8, 9, and 11. The magnetic vector potential **A** and the flux in the core ϕ_c are assumed to be functions of the current in the inductor winding, with sources of **A** external to the inductor neglected. The flux is assumed concentrated in the core with no leakage. Each turn has the same $-\partial\mathbf{A}/\partial t$ along it, and each turn is effectively a closed loop. Integrating in the direction from *a* to *b* around one turn, the electromotance acting around one turn is found by equation 8 to be

$$e = \oint \left(-\frac{\partial \mathbf{A}}{\partial t}\right) \cdot d\mathbf{l} = -\frac{d\phi_c}{dt}$$

The electromotance from *a* to *b* is the electromotance of the *n* turns in series and is

$$e_{ab} = ne = -n\frac{d\phi_c}{dt}$$

The *Ri* product from *a* to *b* obviously is the resistance of the winding from *a* to *b*, times the current in the winding, or

$$(Ri)_{ab} = R_{ab}i_{ab}$$

Equation 11 shows that the potential decrease from terminal *b* to terminal *a* is

$$v_{ba} = e_{ab} - (Ri)_{ab} = -n\frac{d\phi_c}{dt} - R_{ab}i_{ab} \qquad (17)$$

Equation 17 is a relation between the terminal potential difference and the current in the winding, since ϕ_c is a function of i_{ab}.

Although neither term on the right of equation 17 is a potential difference, the potential difference of the terminals is correctly obtained from the equivalent

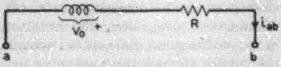


Fig. 5. Equivalent circuit of the inductor of Fig. 3

$$v_0=e_{ab}=-n\frac{d\phi_o}{dt}$$

$$v_{ba}=v_0-Ri_{ab}$$

circuit of Fig. 5, in which the terms of equation 17 are represented by potential differences. This circuit is useful in connection with the potential difference form of Kirchhoff's voltage law, equation 16, in the analyses of circuits containing the inductor, since the effect of the inductor on the external circuit is completely determined by the terminal potential difference.

An equation similar to equation 17 for the inductor may be obtained for circuits where any type of electromotive component is localized. In every circuit of this type the action of the electromotive component is to produce an electrostatic component which acts on the portion of the circuit outside of the region of the electromotive component.

The concept of power flow and energy interchange is enhanced by the writing of equation 17 in the form

$$e_{ab}i_{ab}=R_{ab}i_{ab}^2+v_{ba}i_{ab} \tag{18}$$

Equation 18 shows that the electromotance e_{ab} may be considered as a measure of the power per unit of current passing from the magnetic field to the circuit. This concept is of importance in electric machinery, where the product of electromotance and current in a d-c motor gives the power being converted to mechanical form.

Example 2

The problem discussed in the introduction may now be solved without hesitation with the aid of the concepts presented above. No information is given concerning the distribution of the sources of the magnetic field, so that the electromotive component cannot be evaluated, equation 11 cannot be solved for v, and so the potential difference form of Kirchhoff's second law, equation 16, cannot be used. However, the induced electromotance around a closed loop is expressed by equation 13 in terms of the magnetic flux linking the loop and may be used in the electromotance form of Kirchhoff's second law, equation 15. For a closed path in the conducting ring, equation 15 gives

$$Ri=-\frac{d\phi}{dt}$$

Equation 15 for the loop composed of segment *ab* of the conducting ring and the adjacent voltmeter leads is

$$R_{ab}i-R_{m1}i_{m1}=0; \quad i_{m1}<<i$$

since no flux links this loop. The discussion of voltmeter readings above has shown that $R_{m1}i_{m1}$ is the indication M_1 of the voltmeter. Since R_{ab} is one-third of the total resistance R, the voltmeter indication is

$$M_1=\frac{1}{3}Ri=-\frac{1}{3}\frac{d\phi}{dt}$$

Thus the analysis of this problem requires no consideration at all of potential differences. Indeed, if the sources of the magnetic field are axially symmetrical with respect to the ring, and if the voltmeters are removed, there will be no separation of charge at all, no potential difference, and no electrostatic component. The electromotive component is uniformly distributed around the ring. In this case, energy from the sources of the magnetic field is applied to the heating of the conducting ring by the electromotive component directly, and the intermediate process of setting up a potential difference, as in the first example, does not exist.

The fact that no investigation of the electrostatic component is necessary does not mean that the electrostatic component plays an unimportant part in the mechanism of electromagnetic induction. Indeed, a complete understanding of the induction process is impossible without considering the role of the electrostatic component.

In general, the circuit illustrated in Fig. 1 will have a conducting ring that is nonuniform along its length, and the sources of the magnetic field will be unsymmetrically placed. The total electric intensity in the conducting ring must satisfy Ohm's law, $\mathbf{J}=\sigma\mathbf{E}$, and the current density $\mathbf{J}$ must vary so that the current in every cross section of the conductor is the same. On the other hand, the distribution of the electromotive component is determined by the positions of the sources of the magnetic field. The function of the electrostatic component, then, is to make up the difference between the total electric intensity and the electromotive component. If at any time the electrostatic component does not have the proper value everywhere, there will be a temporary flow of current in such a manner as to redistribute the charge on the surface of the conducting ring until the proper $\mathbf{E}_s$ is obtained. In a good conductor, this readjustment of the charges occurs so rapidly[6] that it may be considered instantaneous in near-zone circuit analysis.

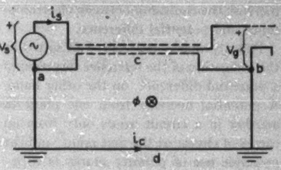


Fig. 6. A ground-loop circuit

Example 3

The necessity of using the electromotance form rather than the potential difference form of Kirchhoff's second law arises in a minority of situations, and the preceding problem is a rather artificial example of such a situation. There are, however, important problems the preferred analysis of which requires the use of the electromotance form of the law, equation 15. The analysis of the squirrel-cage rotor is one. Another, from the medical electronics field, where microvolt signals in the power-frequency band are encountered, is presented in the following.

In Fig. 6 is shown a circuit carrying a circulating current i_c in a ground loop, *acbd*, caused by excessive chassis grounding. The ground loop is linked by 60-cycle magnetic flux from the power transformer and heater leads. It is desired to know what effect the circulating current will have on the circuit.

The electromotive component of electric intensity existing in the various individual parts of the ground loop being practically impossible to determine, the electromotance form of Kirchhoff's second law, which is applicable to closed loops, is used. The Ri products at the tube terminals and the generator are assumed to be effectively equal to the potential differences v_g and v_s at those points. The area of the loop formed by the signal generator, the coaxial cable, and the tube is effectively zero, and so there is no flux linkage in that loop. Using equation 15 clockwise around that loop

$$\Sigma Ri=\Sigma e$$

$$-v_s+v_g+R_{acb}(i_c+i_s)=0$$

or, assuming $i_s<<i_c$

$$v_g=v_s-R_{acb}i_c$$

Going clockwise around the ground loop, *acbd*

$$-(R_{acb}+R_{adb})i_c=-\frac{d\phi_c}{dt}$$

Then

$$v_g = v_s - \frac{R_{acb}}{R_{acb}+R_{adb}} \frac{d\phi_c}{dt}$$

This equation indicates the effects of ground loop orientation and area and of the resistances of the various branches that make up the ground loop.

References

1. ELECTROMAGNETIC INDUCTION, George I. Cohn. *Electrical Engineering*, vol. 68, May 1949, pp. 441–47.

2. ANALYSIS OF ALTERNATING CURRENT CIRCUITS (book), W. R. LePage. McGraw-Hill Book Company, Inc., New York, N. Y., 1952.

3. GENERAL NETWORK ANALYSIS (book), W. R. LePage, S. Seely. McGraw-Hill Book Company, Inc., New York, N. Y., 1952.

4. GESAMMELTE ABHANDLUNGEN (book), G. Kirchhoff. Johann Ambrosius Barth, Leipsig, Germany, 1882, pp. 15–17.

5. UBER DEN DEURCHGANG EINES ELEKTRISCHEN STROMMES DURCH EINE EBENE IN DER INSDESONDERE DURCH EINE KREISFORMIGE VOM STUDIOSUS KIRCHHOFF, H. C. Poggendorff. *Annalen der Physik*, Leipsig, Germany, vol. 64, 1845, p. 513.

6. ELECTROMAGNETIC THEORY (book), J. A. Stratton. McGraw-Hill Book Company, Inc., New York, N. Y., 1951, p. 15.

◆

Discussion

D. R. Mack (General Electric Company, Pittsfield, Mass.): The authors emphasized correctly that electromotance and *Ri* product are not potential differences. Equally important is the fact that the corresponding components of field intensity, $\mathbf{E}_m$ and $\mathbf{E}_s$ are not gradients of potentials. The electromotive component $\mathbf{E}_m$ is produced by a time-varying magnetic field which is nonconservative. Therefore, the electromotance (emf) between two points in the field depends upon the path of integration taken between the points.

Consequently, the value of $\mathbf{E}_m$ at any point is not unique, but depends upon which of the paths of integration through the point is being taken. An example is the field in the gap of an electromagnet, shown in Fig. 1 of the discussion.

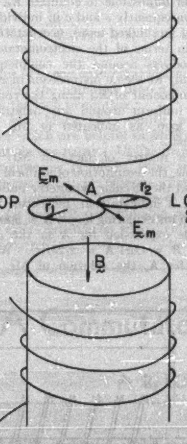


Fig. 1. Field in the gap of an electromagnet

Assume that the flux density **B** is uniform in the gap. If a conducting loop of wire, such as loop 1, is inserted in the gap, the voltage induced in it will be

$$v = \frac{d\phi}{dt} = \pi r_1^2 \frac{d\mathbf{B}}{dt} \tag{1}$$

The electric intensity at any point, such as *A*, around the loop will be

$$\mathbf{E}_m = \mathbf{E} = \frac{v}{2\pi r_1} = \frac{r_1}{2}\frac{d\mathbf{B}}{dt} \tag{2}$$

If loop 1 is replaced by loop 2 (tangent at point *A*, where loop 1 was), the electric intensity at point *A* is

$$\mathbf{E}_m = \frac{r_2}{2}\frac{d\mathbf{B}}{dt} \tag{3}$$

which is different in magnitude and opposite in direction from the value at the same point when computed in loop 1. The value of $\mathbf{E}_m$ at any point depends upon the path it is in.

For this reason, equation 5 of the paper must be interpreted with caution. This equation, and the text just preceding it, indicate that $\mathbf{E}_m$, the force on a unit charge due to nearby changing currents, has at any point a unique value given by equation 5. This implication is not true; equation 5 is meaningful only when it is integrated to find the emf, as the authors have done in equations 8 and 13.

Note in equation 13 that after Stoke's theorem is applied, the curl of $\partial\mathbf{A}/\partial t$, not $\partial\mathbf{A}/\partial t$ itself, is integrated over a surface to obtain the emf. That is

$$e_\Gamma = \iint\left(-\nabla\times\frac{\partial\mathbf{A}}{\partial t}\right)\cdot d\mathbf{s} = \iint -\frac{\partial}{\partial t}(\nabla\times\mathbf{A})\cdot d\mathbf{s} = \iint -\frac{\partial\mathbf{B}}{\partial t}\cdot d\mathbf{s} = -\frac{d\phi_s}{dt} \tag{4}$$

The value of $\mathbf{E}_m$ given by equation 5 of the paper is not the only one which will satisfy equation 4 of the discussion. If any gradient $\mathbf{E}_1$ is added to $-\partial\mathbf{A}/\partial t$ we have

$$\nabla\times\mathbf{E}_m = \nabla\times\left(\mathbf{E}_1 - \frac{\partial\mathbf{A}}{\partial t}\right) = \nabla\times\mathbf{E}_1 - \nabla\times\frac{\partial\mathbf{A}}{\partial t} = -\nabla\times\frac{\partial\mathbf{A}}{\partial t}$$

Thus an $\mathbf{E}_m$ consisting of $-\partial\mathbf{A}/\partial t$ and any gradient $\mathbf{E}_1$ will integrate to the correct value of emf. The potential having the gradient $\mathbf{E}_1$ may be caused by the time-varying magnetic field, and not be present at all under magnetostatic conditions. H. K. Farr (General Electric Company, Pittsfield, Mass.) has suggested that such a potential might explain the anomaly of Fig. 1 of the paper. Charges on the surface of the conductor may cause a potential whose gradient $\mathbf{E}_1$ reconciles the difference between $\mathbf{E}_m$ given by equation 5 of the paper, and the values obtained in equations 2 and 3 of this discussion.

J. J. Smith (General Electric Company, Schenectady, N. Y.): It may be helpful to point out that in some cases it is possible to measure directly the voltages considered by the authors. For instance, in example 1, Fig. 4, of the paper, if another winding of *n* turns is wound over the present exciting winding, the voltage at the terminals of this winding is their electromotance. The voltage across the present exciting winding might be called the "total voltage," which would appear to be a more suitable name than the term "*Ri* product," used by the authors. If these two windings are connected together at one end, then the voltage across the other end is the authors' "potential difference." This latter is usually called the "*Ri* drop." Such an additional coil is used in the Epstein frame for testing magnetic-core loss. Its use allows the value of the magnetic flux density *B* in the sample to be measured directly, and no allowance has to be made for the *Ri* drop, which would have to be done if the total voltage, or voltage across the exciting winding, were the only one measured.

Example 2, Fig. 1, of the paper may be similarly handled. Let another conductor $a'b'$ be laid as close as possible to ab. If a and a' are connected together, then the voltage measured between b and b' when these points are close together is the authors' potential difference v_{ab} between a and b, which again is usually called the *Ri* drop. The important point in making this measurement is that the area of the loop $baa'b'$ should approach zero, so that no voltage can be induced in it due to changing magnetic flux. These measurements can be made as the points a and b approach each other, and thus $\mathbf{E}_s$ can be defined to be v_{ab}/ab, as ab approaches zero. The change of flux in the ring may be measured by placing a second ring as close as possible to the present ring. If this second ring is open at one point, then the voltage across its terminals is their electromotance. However, a complete circuit is always needed for such a measurement and thus it is not easy to go from this measurement to a value of $\mathbf{E}_m$. It can be done by introducing the concept of vector potential along lines previously outlined by the writer,[1] but this is too lengthy to go into here. A simpler procedure is given in the following.

The authors give the well-known definition for electric field strength **E** in equation 2 of the paper as the force on a unit charge. In equations 3 and 5 of the paper, they define $\mathbf{E}_s$ and $\mathbf{E}_m$ by mathematical formula. It would seem preferable to define either $\mathbf{E}_s$ or $\mathbf{E}_m$ in terms of a measurement which can be made similar to the one in equation 2, if possible. We shall do this here, using a definition of electric displacement density. This definition is a slight modification of one being developed by the American Standards Association Committee *C61* on Electrical and Magnetic Units and Magnitudes. Take two small plates close together and connect them together with a wire. Then if they are brought up from a location where the electric field is zero to a point at which there is an electrostatic field, a current will flow in the wire corresponding to the separation of charge on the plates. Call the integral of this current, with respect to time, the total charge Q on the plates; then the electric displacement density **D** is given by

$$\mathbf{D}=\underset{A\to 0}{\text{limit}}\frac{Q}{A}=\underset{A\to 0}{\text{limit}}\frac{1}{A}\int_0^t Idt \qquad (1)$$

when the direction of the plates is such as to make the integral a maximum.

If there is a changing magnetic field present, this will also produce a separation of charge, and thus an additional current in the wire. To separate this current due to the magnetic field from the current due to the electric field, connect the plates together by two wires on opposite edges of the plates. Then the current in these wires due to the electric field will be in the same direction, while the current due to the magnetic field will be in opposite directions. Thus by using a current transformer or oscillograph, or other suitable measuring device, we can measure the integral of the sum of these currents I_1 and I_2 with respect to time. From this, a value of D can be obtained which is independent of the effect of the magnetic field as

$$\mathbf{D}=\underset{A\to 0}{\text{limit}}\frac{1}{A}\int_0^t (I_1+I_2)dt \qquad (2)$$

Thus two separate measurements can be made at any point:

1. The force per unit charge given by equation 2 of the paper;
2. The charge separation per unit area of the two plates given by equation 2 of the discussion.

In general, it is not to be expected that these two measurements should even have the same direction. However, we note that in the measurement of **D** any effect due to a changing magnetic field has been excluded. Thus if the $\mathbf{E}_m$ of the paper is the contribution due to the changing magnetic field, we might expect that **D** would be proportional to $\mathbf{E}_s$ of the paper and write

$$\mathbf{D}=\epsilon\mathbf{E}_s \qquad (3)$$

where ϵ is the permittivity, which is a known quantity for a given medium. The use of **D** in equation 2 of the discussion will also be found very helpful in the more general case involving capacitors, which the authors refer to only briefly.

The objection may be raised that equation 3 of the discussion is generally written $\mathbf{D}=\epsilon\mathbf{E}$, where **E** is the total electric field intensity. There are other reasons for believing that this equation should be revised, but they cannot be gone into here; it is the one given in the older definitions, such as the American Standard Definitions of Electrical Terms.[2] For those who prefer to use the older definitions unchanged, the foregoing discussion would still hold, by changing the name electric displacement density to another name, and giving it a new symbol. The reason for not doing this in the present discussion is that the terminology used here follows some of the newer definitions, such as the American Standards Association *C61* Committee, and the electrical definitions proposed by the International Standards Organization.

Applying the foregoing reasoning to the example 2, Fig. 1, of the paper, the value of $\mathbf{E}_s=\mathbf{D}/\epsilon$ can be measured with two plates at any point along any path from a to b. If the path is a circle close to the ring, it is found that **D** is in the same direction as the ring itself at all points, and its magnitude is also the same at all points. Thus, from such measurements, the integral of $\mathbf{E}_s$ along the path ab can be formed to give the potential difference between these points. The value of the integral is found to be independent of the path in going from a to b. If the integration is around any complete path starting at a and ending at the same point a, then the value of the integral is zero, and thus $\mathbf{E}_s$ is a conservative field.

The total electric field strength $\mathbf{E}_t$ at any point can be measured by the force on a unit charge placed at the point. Subtracting $\mathbf{E}_s$ from $\mathbf{E}_t$ we obtain $\mathbf{E}_m$. However, it is found the integral of $\mathbf{E}_m$ around any closed path starting from b, and passing through a and returning to b, is equal to the change of the flux linking this path. From this it follows that the integral of $\mathbf{E}_m$ from a to b depends on the path. Thus, it does not have a unique value until the path of integration is specified, and hence $\mathbf{E}_m$ is a nonconservative field.

The following suggestion is made with reference to the terminology used by the authors. The integral of $\mathbf{E}_s=\mathbf{D}/\epsilon$ will be called the potential difference; the integral of $\mathbf{E}_t=\mathbf{F}/Q$ will be called the total voltage. The use of "*Ri* product," suggested by the authors for this would produce confusion with existing practice. Call the integral of $\mathbf{E}_m=\mathbf{E}_t-\mathbf{E}_s$ the electromotance. Then all these quantities are measured in volts. Or, it might be said that they are "voltages," as an overall term. It would seem better to reserve the terms "potential difference" and "electromotance" for the two components, conservative and nonconservative, of the total voltage, as discussed in the foregoing. An advantage of this terminology is that it would still hold when capacitors are considered.

REFERENCES

1. THE DERIVATION OF VECTOR POTENTIAL FROM TABLES OF SCALAR POTENTIAL, J. J. Smith. *AIEE Transactions*, vol. 71, pt. I, April 1952, pp. 169–74.

2. AMERICAN STANDARD DEFINITIONS OF ELECTRICAL TERMS. *Publication No. C42—1941*, American Standards Association, New York, N. Y., 1941.

D. W. Spence and **C. R. Cahn**: Mr. Mack raises the timely question of uniqueness in connection with our definitions. Slepian and Cullwick debated this matter somewhat inconclusively[1–6] in *Electrical Engineering* a few years ago. Certainly there are, for instance, several definitions of **A** which satisfy curl **A** = **B**. And if the sources of an **E** field are not known, the authors are aware of no tests which will uniquely separate the $\mathbf{E}_s$ and $\mathbf{E}_m$ components. Making such a separation by test is not the objective, however. Generally speaking, the charge and current distributions, or at least the nature of the charge and current distributions, are known in the apparatus being analyzed. Under such circumstances, it is the authors' experience that the definitions presented, based on an assumed knowledge of charge and current distributions, are powerful aids in the solution of engineering problems and are the most reasonable quantitative definitions of some important electric quantities.

Mr. Mack claims, however, that our admittedly arbitrary definitions do not give unique values of the defined quantities. With this we must disagree. Using the concepts of electromagnetic field theory, our definitions of $\mathbf{E}_s$ and induced $\mathbf{E}_m$ are equivalent to

$$\mathbf{E}_s=-\text{grad } v \qquad (1)$$

$$\text{Induced } \mathbf{E}_m=-\partial\mathbf{A}/\partial t \qquad (2)$$

where v, the scalar electric potential, and **A**, the vector magnetic potential, are defined as follows

$$\text{curl } \mathbf{A}=\mathbf{B} \qquad (3)$$

$$\text{div } \mathbf{A}=-\mu_0\epsilon_0\partial v/\partial t \qquad (4)$$

$$-\text{grad } v=\mathbf{E}+\partial\mathbf{A}/\partial t \qquad (5)$$

These equations, though arbitrarily chosen, are sufficient to determine uniquely the two components of **E**.

They result in **A** being a function only of currents and v a function only of charges.

Concerning his example, it is one of the points of the paper that it usually is difficult or impossible to evaluate $\mathbf{E}_m$ and $\mathbf{E}_s$, and consequently e and v in individual sections of flux-linked loops, necessitating solutions in terms of the electromotances and *Ri* products around the complete loops. Certainly we see no justification in assuming the component of $\mathbf{E}_m$ along the conductors to be uniform around any arbitrarily located loop, as indicated in Mr. Mack's equation 2.

As a matter of fact, in Mr. Mack's example, the assumption of uniform **B** in the gap and the circular symmetry indicated in his Fig. 1 indicate that, neglecting the currents in his exploring loops, $\mathbf{A}=\theta_1 krf(t)$ is a possible expression for **A** in the air gap, giving $\mathbf{B}=\text{curl }\mathbf{A}=\mathbf{z}_1 2kf(t)$. With this value for **A**, the solution of Mr. Mack's

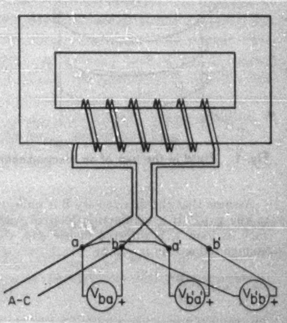


Fig. 2. Apparatus suggested by J. J. Smith

example will be undertaken. The authors are aware that the assumption of an exactly uniform, time-varying **B** field is barred by Maxwell's equations. It is assumed, however, that the frequency is low enough that the assumption of uniformity of **B** is a good approximation.

Mr. Mack's term "voltage" and symbol v are not defined and apparently differ from the definitions presented in our paper. In our terminology, his equation 1 becomes

$$e=-\frac{d\phi}{dt}=-\pi r_1^2\frac{\partial \mathbf{B}}{\partial t}=-2\pi r_1^2 k\frac{df(t)}{dt} \qquad (6)$$

where e is the emf acting around loop 1. His equation 2 is replaced by two equations

$$\mathbf{E}_m=-\frac{\partial \mathbf{A}}{\partial t}=-\theta_1 kr\frac{df(t)}{dt} \qquad (7)$$

$$\mathbf{E}=\rho\mathbf{J}=\mathbf{l}_1\frac{RS}{l}\frac{i}{S}=\mathbf{l}_1\frac{Ri}{2\pi r_1}=\mathbf{l}_1\frac{e}{2\pi r_1}$$
$$=\mathbf{l}_1\left(-\frac{r_1}{2}\frac{\partial \mathbf{B}}{\partial t}\right)=\mathbf{l}_1\left(-r_1 k\frac{df(t)}{dt}\right) \qquad (8)$$

where Ri is the Ri product around the loop and is equal to e by equation 15 of the paper, and $\mathbf{l}_1$ is a unit vector tangential to the loop at each point on the loop. $\mathbf{E}_m$ is independent of what loop is under consideration, so that we do not get Mr. Mack's equation 3.

As Mr. Farr and Mr. Mack suggest, there is an electrostatic component $\mathbf{E}_s$ arising from charge separation on any loop not coaxially placed within the gap. Its value along or within one of the loops, where **E** is known, is

$$\mathbf{E}_s=\mathbf{E}-\mathbf{E}_m=k\frac{df(t)}{dt}\{\theta_1 r-\mathbf{l}_1 r_1\} \qquad (9)$$

This value of $\mathbf{E}_s$ can be integrated around the loop to find the potential drop v from one point on the loop to another.

Like Mr. Mack, Dr. Smith presents examples which are helpful in bringing out the points in our paper. Unlike many electrical engineering problems, these are such that it is not intuitively obvious what "voltages" are present. The values of current present are uniquely determinable, but what voltage exists depends on what "voltage" is defined to be. From this standpoint, it is unfortunate that Dr. Smith also uses the word "voltage" without defining it.

More seriously, as nearly as the authors can ascertain Dr. Smith's meaning, he seems to have misinterpreted two of our definitions and consequently to have come to some incorrect conclusions.

In Fig. 2 of the discussion, the leads are brought out to a region remote from the core where the $\mathbf{E}_m$ component is negligible, so that $\mathbf{E}=\mathbf{E}_s$. As discussed in the paper, the voltmeters read the Ri product through themselves, that is, the integral of **E** through themselves. But since $\mathbf{E}=\mathbf{E}_s$, this is the same as the potential difference, so the voltmeters also read the potential difference between the terminals to which they are connected. This is usually the case in electric apparatus, though not always. It will now be assumed that the readings of these voltmeters are what Dr. Smith refers to when he speaks of "voltage."

If e_{ab} is the electromotance acting in winding a', $(Ri)_{ab}$ the Ri product through winding ab, and $e_{a'b'}$ and $(Ri)_{a'b'}$ are the corresponding quantities in winding $a'b'$, then, by equation 11 of the paper

$$v_{b'a'}=e_{a'b'}-R_{a'b'}i_{a'b'} \qquad (10)$$

$$v_{ba}=e_{ab}-R_{ab}i_{ab} \qquad (11)$$

$$v_{b'b}=e_{a'b'}-R_{a'b'}i_{a'b'}-e_{ab}+R_{ab}i_{ab} \qquad (12)$$

The windings are such that $e_{a'b'}=e_{ab}$, and it is assumed that the voltmeters are high resistance so that $i_{a'b'}=0$. Then the voltage referred to in Dr. Smith's second sentence is $V_{b'a'}=e_{a'b'}=e_{ab}$, the electromotance, just as he says. The voltage of his third sentence is $v_{ba}=e_{ab}-R_{ab}i_{ab}$, which cannot be described by one term except as the potential drop from b to a, v_{ba}. The authors had no intention of indicating that this quantity is the Ri product. The voltage of Dr. Smith's fourth sentence is $v_{b'b}=e_{ab}-R_{a'b'}(0)-e_{ab}+R_{ab}i_{ab}=R_{ab}i_{ab}$, the Ri product in the exciting winding, and is not the potential difference across either winding. Dr. Smith's term "total voltage," if substituted for the authors' "Ri product," is logical in that it is the integral of total, or net, **E**. There is some question whether it is desirable from a standpoint of connotation, however.

The same reasoning applied to Dr. Smith's second paragraph seems to invalidate his determination of $\mathbf{E}_s$. His limit v_{ab}/ab gives **E**, where his v_{ab} refers to voltmeter reading, not potential difference.

Dr. Smith's "D-Measurer" is very interesting. Dr. Slepian argues,[2] however, that there is no test which will separate the $\mathbf{E}_s$ and $\mathbf{E}_m$ components of **E**, and we cannot see where the **D** Dr. Smith measures will be other than $\epsilon\mathbf{E}$ ($=\epsilon\mathbf{E}_s-\epsilon\partial\mathbf{A}/\partial t$, in electromagnetic fields). The $-\partial\mathbf{A}/\partial t$ component of electric intensity should charge the test capacitor just as well as the $\mathbf{E}_s$ component does. True, any rotational component of the $-\partial\mathbf{A}/\partial t$ field, as measured by $-d\mathbf{B}/dt$, and as evidenced by an increase or decrease of $-\partial\mathbf{A}/\partial t$ from one side of the capacitor to the other, is averaged out, but $-d\phi/dt$ is not the only way that changing currents, that is, magnetic effects, affect apparatus such as Dr. Smith's search capacitor. The authors suggest in the paper that $-\partial\mathbf{A}/\partial t$ is a more productive place to start the analysis of such out-of-the-ordinary apparatus than is $-d\phi/dt$.

In summary, the authors feel that the concepts presented in the paper are very useful in reasoning about the operation of much typical electric apparatus. They provide a common method of analyzing electrochemical, electrothermal, and perhaps piezoelectric, as well as electromagnetic, apparatus. But in purely electromagnetic situations, the resolutions of the **E** field has conceptual significance only, and is not subject to unique experimental verification.

REFERENCES

1. ELECTROSTATIC OR ELECTROMAGNETICALLY INDUCED ELECTRIC FIELD?, J. Slepian. *Electrical Engineering*, vol. 68, Oct. 1949, p. 877.

2. *Ibid.*, Nov. 1949, pp. 985–87. Note apparent error in equation 17, p. 986; it should probably read: div $\mathbf{A}+\frac{1}{c}\frac{\partial\psi_L}{\partial t}=0$.

3. *Ibid.*, E. G. Cullwick, vol. 69, Feb. 1950, pp. 187–88. Note apparent errors in equations 7, 8, and 10, p. 188, where $\mathbf{E}_s$ should probably be replaced by **E**.

4. *Ibid.*, J. Slepian, p. 188.

5. *Ibid.*, E. G. Cullwick, May 1950, p. 481.

6. *Ibid.*, J. Slepian, p. 481.

A Commutatorless D-C Tachometer

A. R. ECKELS
ASSOCIATE MEMBER AIEE

W. R. PECK
NONMEMBER AIEE

IN MANY applications the commutator noise produced by conventional tachometers presents great problems. This is particularly true in cases where it is desirable to measure speed variations which have the same frequency components as the commutator noise. It is also true in some servomechanism systems where the commutator noise will saturate the amplifiers or will introduce a spurious response.

Several devices have been developed to avoid these difficulties and are described in recent literature.[1,2] The tachometer described here is another approach toward the problem of constructing a tachometer with a small noise-to-signal ratio which has the required linear response. The operation of this tachometer is based upon the familiar current balance principle.

Description

The four principal elements of the tachometer are shown in Fig. 1. Rotation of the magnet mounted on the input shaft induces eddy currents in the disk, which in turn produce a torque on the disk proportional to the speed of rotation. This torque will tend to move the shaft holding the disk mirror, and coil. A movement of the mirror will deflect

Paper 53-292, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 16, 1953; made available for printing May 4, 1953.

A. R. ECKELS and W. R. PECK are with North Carolina State College, Raleigh, N. C.

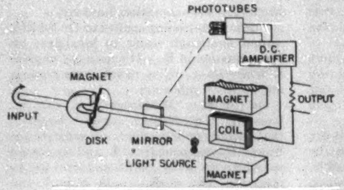


Fig. 1 (left). Pictorial diagram of commutatorless tachometer

Fig. 3 (right). The completed commutatorless d-c tachometer

the beam of light and will cause one phototube to receive a larger proportion of the light. This will cause an unbalance between the currents of the phototubes, thereby producing a voltage signal which is then amplified. The output of the amplifier is impressed upon the coil of a D'Arsonval-type instrument and a series resistance. Thus the deflection of the mirror produces a current in the coil. This current will exert a restoring torque on the coil in such a direction as to oppose the original movement of the shaft, and of such magnitude as to balance the torque produced on the disk by the rotation of the magnet. Since the current in the coil also flows through the resistance, the voltage drop across this resistance will be proportional to the speed of the input shaft.

The magnet and disk assembly is identical to one employed in a commercial indicator tachometer.[3] This element has the important function of exerting a torque on the disk shaft which is a linear function of the input angular speed.[4]

A beam of light and phototube system is employed for the shaft position detector. The maximum signal occurs for a 1-degree deflection of the mirror. Other standard types of position detectors could have been employed.

The direct-coupled amplifier, including the two phototubes, is shown schematically in Fig. 2. Differential amplifier stages are employed throughout in order to minimize drift and to reduce the effect of variations in the power supply voltage. The output stage consists of four *6L6*'s paralleled two by two, and gives a source impedance of about 650 ohms, with a current capacity of ±60 milliamperes. Each stage of voltage amplification has a balance arrangement, and the gain control is the variable resistor shunting the grids of the second *6SL7*.

The D'Arsonval movement is from a 30-milliampere meter, the coil frame of which was altered to eliminate the damping effect of the frame. The torque of this coil is a linear function of the coil current. The maximum current produces a signal of ±1.2 volts across the resistance in series with the coil.

The completed tachometer shown in Fig. 3 consists of three separate units. The first contains the d-c power supply and two separate a-c heater supplies. The second unit contains the input shaft, the disk, coil, and mirror assembly, the beam of light, and phototube system, and the first stage of voltage amplification. The third unit contains the amplifier with the exception of the first amplification stage.

Fundamental Concepts

The tachometer will now be analyzed to determine the way in which its accuracy and frequency response depend on the constants of the closed loop system. For convenience of analysis the elements shown in the pictorial diagram of Fig. 1 are represented by the block diagram of Fig. 4. In this diagram Q_{in} represents the torque exerted on the disk by the rotation of the input shaft, while Q_c represents the electrical torque produced by the current flowing in the coil. Q_s is the net torque exerted on the moving system consisting of the disk, mirror, and coil. Thus

$$Q_s = Q_{in} - Q_c \qquad (1)$$

However, Q_c will be related to Q_s by some frequency dependent function which is called the loop gain and is denoted as $K_1G_1(jw)$, or

$$Q_c(jw) = [K_1G_1(jw)]Q_s(jw) \qquad (2)$$

By eliminating Q_s from equations 1 and 2, it can be shown that

$$\frac{Q_c(jw)}{Q_{in}(jw)} = \frac{K_1G_1(jw)}{1+K_1G_1(jw)} \qquad (3)$$

In reference 4 it is shown that

$$Q_{in} = K_2N_{in} \qquad (4)$$

and the characteristics of a D'Arsonval instrument movement and Ohm's law require that

$$Q_c = K_3E_R \qquad (5)$$

where E_R is the voltage across the resistor

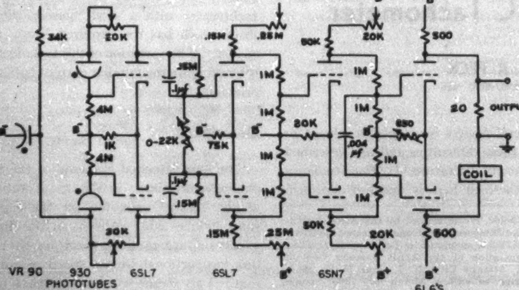


Fig. 2. Schematic diagram of amplifier

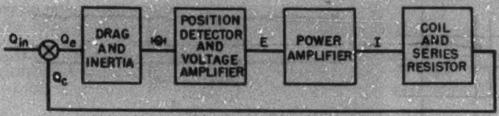


Fig. 4. Block diagram of servo loop

Table I. Noise-to-Signal Ratio of the Commutatorless Tachometer and of a Standard D-C Generator Tachometer

Speed, Rpm	Noise to Signal Ratio	
	Standard Tachometer	Commutatorless Tachometer
120	0.4	0.046
240	0.5	0.029
450	0.46	0.020
650	0.36	0.020
800	0.167	0.021
1,000	0.04	0.017
1,100	0.03	0.013

in series with the coil. Substitution of equations 4 and 5 in equation 3 yields

$$\frac{E_R}{N_{in}}=\frac{K_2}{K_3}\left[\frac{K_1G_1(jw)}{1+K_1G_1(jw)}\right] \tag{6}$$

Inspection of equation 6 shows that for signals having frequency components such that

$$K_1G_1(jw)\gg 1 \tag{7}$$

the tachometer will yield an output voltage signal which is proportional to the angular speed of the input shaft. An analytical expression for $K_1G_1(jw)$ may be obtained from the differential equations of the components.[6] It is important to realize, however, that the nature of $K_1G_1(jw)$ is such that special precautions must be taken in order that the system will be stable when $K_1G_1(jw)$ is increased to a suitable level. The problem of obtaining a stable system can be solved by the application of techniques developed for automatic control systems.[6] These techniques include careful design of the mechanical system to minimize the principal time lags and the introduction of compensating networks.

Test Results and Conclusions

The tests carried out on the tachometer included measurements to allow the determinations of $K_1G_1(jw)$ as a function of frequency and three performance tests on the tachometer.

Amplitude and phase loci of $K_1G_1(jw)$ obtained from the test data are shown in Fig. 5. The curves shown with solid lines represent the completed system which includes the resistance-capacitance compensating network in the voltage amplifier. For suitable operation it is necessary to limit the phase of $K_1G_1(jw)$ to approximately 150 degrees at the frequency for which the amplitude $K_1G_1(jw)$ is zero decibels. A comparison of the solid and dashed curves will show that the inclusion of the compensating network made it possible to increase K_1 by 60 decibels and still meet the criterion mentioned previously for suitable operation.

The results of the first performance test, shown in Fig. 6, indicate that the output signal is a linear function of the input angular velocity. So long as the condition stated in equation 7 is met, the calibration of the instrument is dependent only upon K_2 and K_3. As defined in equation 4, K_2 is the constant of the magnet and disk assembly. The particular unit used was designed to be accurate within 10 rpm up to a speed of 3,000 rpm.[3] As defined in equation 5, K_3 is the constant of the D'Arsonval instrument and the output resistor. These components can be designed for very high accuracy.

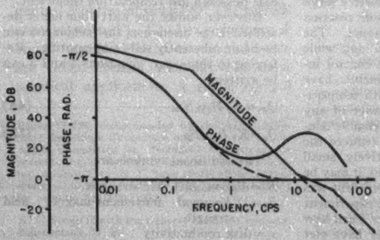
The second performance test determined the output voltage per revolutions per minute as a function of frequency. The input function of angular velocity was obtained by the oscillations of a weight clamped to the input shaft which was acted on by the torsional resilience of a long, slender wire. The frequency could be varied by clamping the wire at various lengths in a stationary vise. Thus the input assumes the form of a sinusoid with easily determined constants. The output signal was recorded by a type *BL202* Brush Development Company magnetic recorder. The calculated values of the maximum output signal voltage divided by the maximum input revolutions per minute as a function of frequency are plotted in Fig. 7. The low-frequency portion of the curve is asymptotic to 0.885 millivolt per rpm. This value is obtained from Fig. 6. Although no data were obtained for signal frequencies in excess of 15 cycles per second, the curve in this region can be expected to approach K_2/K_3 times the solid curve in Fig. 5. This is true since above 15 cycles per second $K_1G_1(jw)$ is less than 1, and equation 6 reduces to

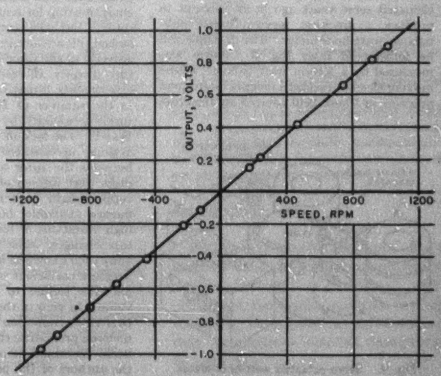
$$\frac{E_R}{N_{in}}\cong\frac{K_2}{K_3}K_1G_1(jw) \tag{8}$$

The final test determined the noise-to-

Fig. 5 (below). Magnitude and phase of $K_1G_1(jw)$ as a function of frequency. Dashed lines indicated uncompensated system

Fig. 6 (right). Output voltage as a function of input angular velocity





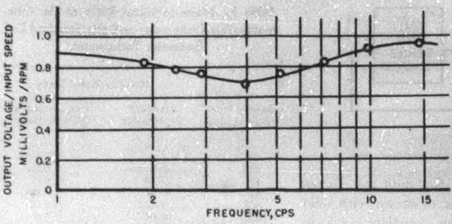


Fig. 7 (above). Ratio of output voltage to input speed as a function of frequency

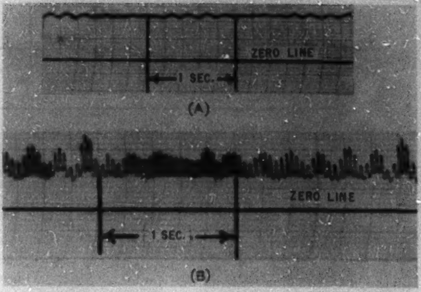


Fig. 8 (right). The output voltage of the commutatorless d-c tachometer and of a normal d-c generator for the same input speed

signal ratio of the commutatorless tachometer and of a normal d-c generator tachometer. Fig. 8 shows sample oscillograms of the output voltages for the commutatorless tachometer and a d-c generator tachometer with the same angular speed input. It should be noted that the low-frequency component of voltage is common to both oscillograms and is unquestionably caused by a variation in the shaft speed. The noise-to-signal ratio for various shaft speeds of both tachometers is tabulated in Table I.

From these tests it can be concluded that the basic principle of operation of this type of tachometer is sound. There would need to be some definite improvements in the mechanical construction of the disk, position detector, and coil assembly to allow a further increase in the frequency response, as the shape of $K_1G_1(jw)$ is primarily determined by this assembly.[5] The ability of the tachometer to respond to significant high-frequency components of speed with reasonable amplifier power is also determined by the D'Arsonval movement. With a D'Arsonval movement having characteristics matching those of the pen motor described in reference 5, a tachometer could be designed which would respond to frequency components of speed ranging from direct current to a frequency of 200 cycles per second.

References

1. Measurement of Speed Variation in Rotating Equipment, **E. G. Manning.** *Bulletin No. 50,* Department of Engineering Research, North Carolina State College, Raleigh, N. C., 1951.

2. A Recording Tachometer for Measuring Instantaneous Angular Speed Variations, **S. P. Bartles.** *Electrical Engineering,* vol. 70, Sept. 1951, pp. 816–19.

3. The Magnetic-Drag Tachometer, **R. G. Ballard.** *AIEE Transactions,* vol. 61, 1942, pp. 366–368.

4. Permanent Magnets in Drag Devices and Torque Transmitting Couplings, **R. J. Parker.** *General Electric Review,* Schenectady, N. Y., vol. 52, 1949, pp. 16–20.

5. A High-Speed Direct-Writing Oscillograph, **I. S. Blumenthal, A. R. Eckels.** *AIEE Transactions,* vol. 71, pt. I, Jan. 1952, pp. 50–56.

6. Automatic Feedback Control (book), **W. R. Ahrendt, J. F. Taplin.** McGraw-Hill Book Company, Inc., New York, N. Y., 1951.

◆

Discussion

Perry A. Borden (The Bristol Company, Waterbury, Conn.): In outlining the development of a null-type magnetic-drag tachometer, the authors have made no reference to one problem which is always present in devices of this class. I refer to the effect of temperature changes on the resistivity of the disk or cup which forms the rotor. I recall certain experiments in which I participated some years ago in an attempt to produce a precision servo-actuated null-balance watt-hour meter. The principle will be understood from Fig. 9. Power was measured by a Kelvin watt balance, whose moving element carried a magnet or group of magnets in whose field rotated an armature in the form of a nonmagnetic metal disk driven by a variable-speed motor mechanism. The speed of the driving mechanism was regulated by the torque developed on the moving element of the wattmeter, with a resultant continual tendency for the rate of rotation of the disk to be proportional to the measured power.

By virtue of the driving force and the retarding influence both being exerted on a common conducting member, the conventional watt-hour meter is to some extent self-compensating for temperature effects on the rotor; but the experimental instrument lacked this advantage, and changes in the resistivity of the rotor material with temperature directly affected its accuracy. These experiments, it may be said, were carried out in the summer of 1914, and were prematurely terminated by the outbreak of World War I. The lack of today's sensitive servo systems necessitated a maximum reaction between the rotor and the magnet. The disks were formed of aluminum, and while copper might have made more control influence available, both these metals have high variations of resistivity with temperature changes. The high resistance of any alloy then known as having a negligible temperature coefficient would have reduced the effective torque to a prohibitively small value. In view of these obstacles, it may be that the termination of the experiment was no more premature than was the experiment itself. It would be of interest to learn how the authors of the present paper have met and overcome this important source of error.

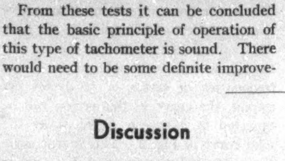


Fig. 9. Servo-actuated watt-hour meter

A. R. Eckels and **W. P. Peck:** The authors wish to express their appreciation for Mr. Borden's comments and for the very interesting description of an early example of the family of devices based upon the current balance principle.

Construction details of the magnet and disk assembly were omitted from this paper, as they are fully described in references 3 and 4 of the paper. The two principal means employed to reduce the sensitivity to temperature changes are to make the disk of an aluminum-manganese alloy which has a low temperature coefficient, and to use temperature-sensitive magnetic shunts in the magnet assembly. The output resistor must in this case possess a low temperature coefficient.

However, unlike the watt-hour meter described in the discussion, the tachometer can be made inherently self-compensating. Referring to the paper, equations 4 and 5 can be written

$$Q_{in}=K_2'B_2\gamma_2N_{im} \qquad (4A)$$

$$Q_c=K_3'B_3\gamma_3E_R \qquad (5A)$$

The additional symbols are

B_2 = disk-magnet field strength
B_3 = D'Arsonval movement-magnet field strength
γ_2 = disk conductivity

γ_3 = output-resistor conductivity

Thus, if the magnetic and conducting elements of each assembly employ similar materials in corresponding parts, equation 6 will become

$$\frac{E_R}{N_{in}} = \frac{K_2'}{K_3'}\left[\frac{K_1G_1(jw)}{1+K_1G_1(jw)}\right]$$

Thus the temperature-sensitive components do not affect the results. This assumes the same ambient temperatures and thermal transient properties for corresponding parts and neglects very small effects such as expansion and contraction.

It should also be noted that where speed variations are being measured, slow changes of calibration are relatively unimportant.

Theory of Magnetic Amplifiers with Square-Loop Core Materials

H. F. STORM
MEMBER AIEE

Synopsis: The purpose of this paper is to explain the operation of magnetic amplifiers and to express power out, power gain, and time constant of the amplifier in terms of core and rectifier properties. The magnetic amplifiers under discussion are of the so-called low-control circuit-impedance type.[1] To reduce the complexity of the analysis, many simplifying assumptions are made. While these assumptions introduce inaccuracies in the equations, the latter still serve effectively in establishing first-order relations among the variables. A large part of the analysis is applicable to core materials whose dynamic hysteresis (B-H) loop is not strictly rectangular, such as grain-oriented silicon steel.

THE FOLLOWING analysis applies to magnetic-amplifier circuits of the center-tap, bridge type, or a-c variety. Such circuits are called amplistats,[2] Magamps,[3] transductors,[4] self-saturated magnetic amplifiers,[5] etc. Fig. 1 shows the circuit diagram of the center-tap amplistat.

Circuit

The secondary *TS* of the supply transformer energizes the gate windings (between *X*1, *X*2) of two transformerlike structures *A* and *B*. From there the gate currents flow through rectifiers $REC_{1,2}$ to the load resistance, and return to the neutral *N* of transformer *TS*. The secondaries between *Y*1, *Y*2 of transformers *A* and *B* form the control circuit and hence are called control windings. In most practical applications, a single control winding is used which embraces both legs of units *A* and *B*. In this case, units *A* and *B* are called a saturable reactor. With the assumptions to be listed, the function of the amplistat is the same for two separate, or one common, control winding.

Nomenclature

A_{Fe} = area of active core cross section, centimeters^2
AT = magnetomotive force (mmf), ampere-turns, instantaneous
AT_c = mmf for zero flux on symmetrical dynamic hysteresis loop, ampere-turns, instantaneous
B = flux density, gausses, instantaneous
B_s = saturation flux density, gausses, instantaneous
$B_{s,1}$ = saturation flux density for reactor, being load-matched and operating at maximum permissible temperature, gausses, instantaneous
D_P = dynamic power gain, seconds^{-1}
E = line-to-neutral voltage of supply transformer secondary, volts, average
E_C = control voltage, volts
$e_{G,A}$, $e_{G,B}$ = gate voltage on saturable reactor *A* or *B* respectively, volts, instantaneous
e_O = (nominal) output voltage, volts, instantaneous
E_O = (nominal) output voltage, volts, average
e_r = rectifier reverse voltage, volts, instantaneous
f = supply frequency, cycles per second
H = component of magnetic intensity produced by winding, oersteds, instantaneous
H_c = H for zero flux on symmetrical, dynamic hysteresis loop, oersteds, instantaneous
i_C = control current, amperes, instantaneous
I_C = control current, amperes, average
$\mathbf{i}_C$ = per-unit control current
$i_{C,r}$ = control current required by rectifier reverse current, amperes, instantaneous
$I_{C,r}$ = control current required by rectifier reverse current, amperes, average
I_E = exciting current, amperes, average
$\mathbf{i}_E$ = per-unit exciting current
$i_{G,A}$, $i_{G,B}$ = gate current in saturable reactor *A* or *B*, respectively, amperes, instantaneous
$i_{G,r}$ = gate current caused by rectifier reverse current, amperes, instantaneous
$i_{G,s}$ = saturation component of gate current, amperes, instantaneous
$i_{G,x}$ = exciting component of gate current, amperes, instantaneous
i_L = load current, amperes, instantaneous
$i_{L,s}$ = saturation component of i_L
$i_{L,x}$ = exciting component of i_L
I_L = load current, amperes, average
$I_{L,s}$ = saturation component of I_L
$I_{L,r}$ = exciting component of I_L
$I_{L,m}$ = maximum load current, amperes, average
i_r = rectifier reverse current, amperes, instantaneous
$I_{r,m}$ = maximum rectifier reverse current, amperes, averaged over 1/2 cycle
$\mathbf{i}_r$ = per-unit rectifier reverse current
k_f = form factor of load current
K_{AT} = ampere-turn gain
$K_{AT,s}$ = saturation component of K_{AT}
$K_{AT,s,0}$ = $K_{AT,s}$ with zero rectifier reverse current
$K_{AT,s,r}$ = $K_{AT,s}$ with rectifier reverse current present
$K_{AT,x}$ = exciting component of K_{AT}
K_E = voltage gain
K_I = current gain
K_P = power gain
l_{Fe} = length of mean path of magnetic circuit, centimeters
N_C = number of turns of one control winding
N_G = number of turns of one gate winding
P_L = useful power in R_L, watts
R_C = control circuit resistance, ohms
$R_{G,A}$, $R_{G,B}$ = gate resistance in saturable reactor *A* or *B*, respectively, ohms
R_L = load resistance, ohms
R_O = nominal output resistance, ohms
R_r = simulated rectifier reverse resistance, ohms
t = time, seconds
T = time constant, seconds
α = starting angle of saturation, radians
κ = see equation 47
μ_{sq} = square-loop permeability, gausses per oersted
ϕ_A, ϕ_B = core flux in saturable reactor *A* or *B* respectively, maxwells, instantaneous
Φ_1 = starting flux, maxwells, instantaneous
Φ_s = saturation flux, maxwells, instantaneous
ω = angular frequency, seconds^{-1}

Assumptions

The control-circuit resistance R_C is assumed small enough to prevent appreciable interference with the a-c components of the control current; the net voltage in the control circuit is so small that the resultant core-flux changes can be neglected.

Paper 53-284, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 16, 1953; made available for printing April 29, 1953.

H. F. Storm is with the General Electric Company, Schenectady, N. Y.

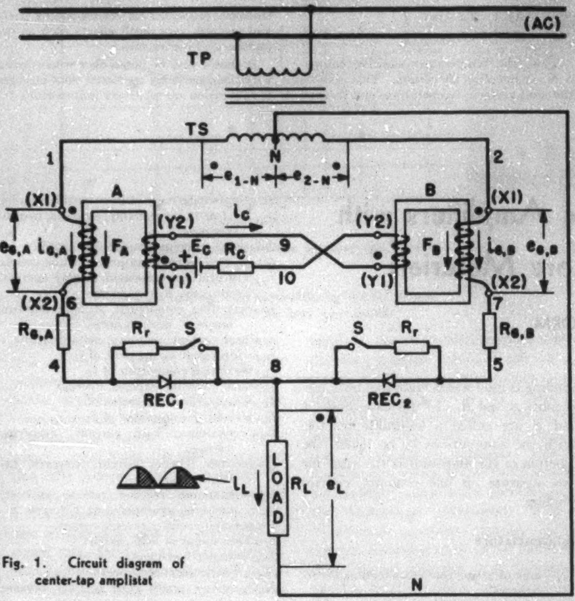


Fig. 1. Circuit diagram of center-tap amplistat

During the exciting interval $0<\omega t<\alpha$ the resistances encountered by the load current are neglected; during the saturation interval ($\alpha<\omega t<\pi$), the reactances encountered by the load current are neglected. The rectifiers have constant forward resistance, and the reverse resistance is considered infinite for the time being.

When referring to the B-H loop, the dynamic loop is always meant, relating flux densities B to exciting ampere-turns iN. This loop is obtained[6-8] by the energization of core A or B with a sinusoidal voltage e_{1-N}

$$e_{1-N}=\frac{\pi}{2}E\sin\omega t \tag{1}$$

and frequency f

$$\omega=2\pi f \tag{2}$$

at which the operation of the amplistat is to be investigated. Assuming a theoretical rectangular B-H loop, as shown in Fig. 2, the magnitude of the applied a-c voltage is just sufficient to cause a sinusoidal flux swing equal to the saturation flux Φ_s

$$E=4f\Phi_s N_G 10^{-8} \tag{3}$$

where

$$\Phi_s=B_s A_{Fe} \tag{3A}$$

When dealing with a real B-H loop, B_s and Φ_s are replaced by the actual amplitudes of flux density and flux.

The exciting ampere-turns iN are usually expressed in the form

$$H=0.4\pi iN/l_{Fe} \tag{4}$$

The actual magnetic intensity in the core is different from that expressed by equation 4 on account of eddy currents.[9] However, for purposes of circuit analysis, it is precisely the relation between flux and exciting ampere-turn which is of interest.

The ampere-turns indicated by the dynamic B-H loop for zero flux are called, for lack of a better term, coercive ampere-turns AT_c, and the magnetic intensity H_c produced by AT_c is called the coercive force

$$H_c=0.4\pi AT_c/l_{Fe} \tag{5}$$

The symmetrical B-H loop is approximated by a rectangle, Fig. 2. The unsymmetrical hysteresis loops[10] are derived by preserving the top and the positions of the sides of the symmetrical B-H loop, and by varying the position of the horizontal bottom of the B-H loop.

Exciting Interval

It is assumed that at $\omega t=0$ the core flux $\phi_A=\Phi_1$ and the core flux $\phi_B=\Phi_s$, point 1, Figs. 2(A,2), 2(B,2). The line-to-neutral voltage e_{1-N} requires an increasing core flux ϕ_A in saturable reactor A

$$\Phi_A=\Phi_1+\int_0^{\omega t}e_{1-N}10^8 N_G^{-1}d(\omega t) \tag{6}$$

for $0\leqslant\omega t\leqslant\alpha$

and substituting from equations 1 and 3

$$\phi_A=\Phi_1+\Phi_s(1-\cos\omega t) \tag{7}$$

for $0\leqslant\omega t\leqslant\alpha$

To make the rising flux ϕ_A possible, the operating point on the B-H loop of core A moves toward the right from 1 to 2, Fig. 2(A,2); this path, being horizontal, indicates that the gate winding A will not exhibit inductance. Since the circuit resistance is assumed to be zero during the exciting interval, the gate current $i_{G,A,x}$ increases abruptly, as illustrated in Fig. 2(A,4). During the same interval, the exciting gate current $i_{G,B,x}$ of saturable reactor B, Fig. 2(B,4) is zero because of the blocking action of rectifier REC_2, Fig. 1, and the absence of any induced voltage from saturable reactor A.

When the current $i_{G,A,x}$ has increased to such a value that its ampere-turns $i_{G,A,x}N_G$ equal the coercive ampere-turns AT_c, the vertical ascending part of the B-H loop, Fig. 2(A,2), is reached at point 2. The time is still $\omega t=0$. The flux ϕ_A becomes ready to increase along the vertical side 3–4 and saturable reactor A can function as a transformer. The gate voltage $e_{G,A}$ being now identical with the supply voltage e_{1-N} is transformed into the control circuit. As a result, control current i_C will flow, moving operating point 2, Fig. 2(B,2), on the B-H loop of saturable reactor B toward 3. Because 2–3 is horizontal, the inductance of the control winding of saturable reactor B is zero and the control ampere-turns $i_C N_C$ increase abruptly until $i_C N_C=-AT_c$, Fig. 2(L,5), assuming negligible control-circuit resistance. To maintain in saturable reactor A net ampere-turns equal to AT_c, exciting gate ampere-turns $i_{G,A,x}N_G$ must increase to $2AT_c$, from point 2 to 3, Fig. 2(A,4).

At point 3 the vertical, descending branch of the B-H loop of core B is reached. Neglecting the voltage drop in the control circuit, gate voltage $e_{G,A}$, transformed into saturable reactor B, causes a core flux ϕ_B.

$$\phi_B=\Phi_s\cos\omega t \tag{8}$$

for $0\leqslant\omega t\leqslant\alpha$

As a further result of this transformation, voltage $e_{G,B}$, Fig. 2(B,1), appears across the terminals of the gate winding of saturable reactor B. It will be seen from equations 7 and 8 that core flux ϕ_B

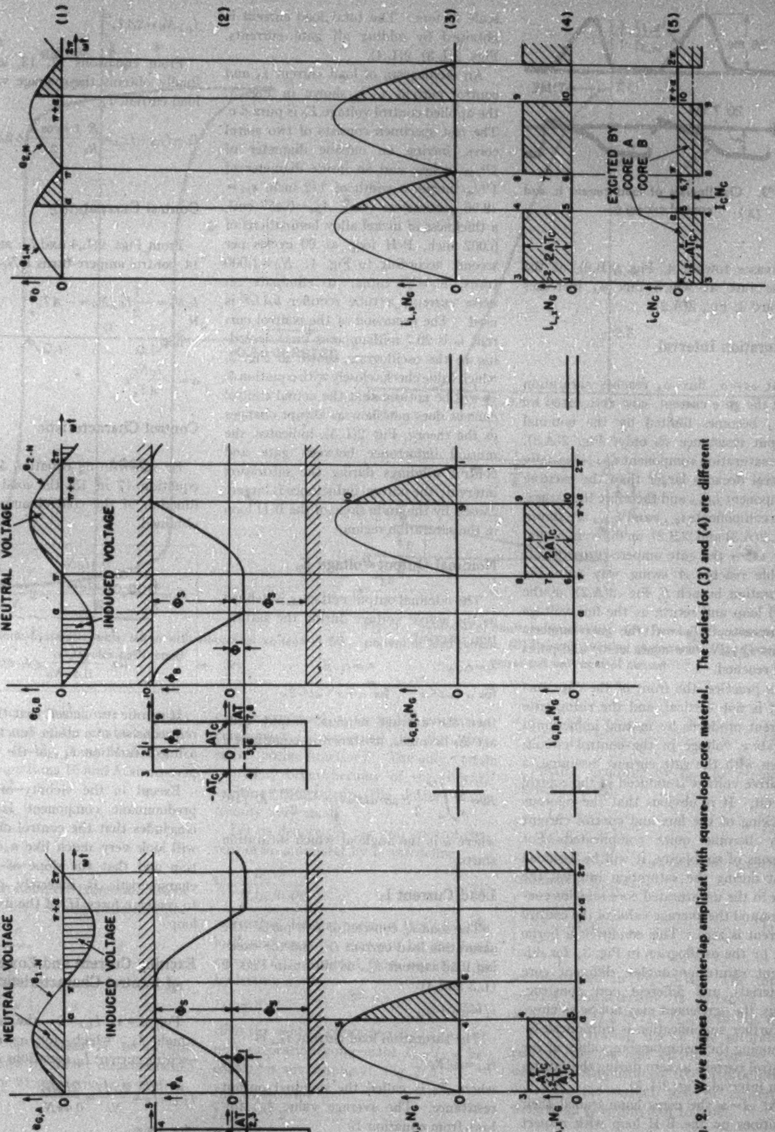

(A)
(B)
(L)
DIRECTLY APPLIED LINE TO NEUTRAL VOLTAGE
INDUCED VOLTAGE
DIRECTLY APPLIED LINE TO NEUTRAL VOLTAGE
INDUCED VOLTAGE
EXCITED BY
CORE A
CORE B
(1)
(2)
(3)
(4)
(5)
$e_{G,A}$
$e_{G,B}$
e_0
$e_{1,N}$
$e_{2,N}$
Φ_A
Φ_B
Φ_S
Φ_1
AT_C
$-AT_C$
AT
$2AT_C$
$i_{G,A,1}N_G$
$i_{G,B,2}N_G$
$i_{L,1}N_G$
$i_{L,2}N_G$
i_CN_C
I_CN_C
ωt
α
π
$\pi+\alpha$
2π
0

Fig. 2. Wave shapes of center-tap amplistat with square-loop core material. The scales for (3) and (4) are different

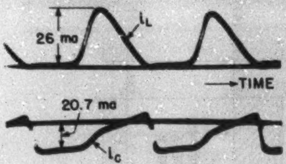


Fig. 3. Oscillogram of load current i_L and control current i_C

decreases toward 4, Fig. 2(B,2), at the same rate at which flux ϕ_A increases toward 4, Fig. 2(A,2).

Saturation Interval

At $\omega t=\alpha$, flux ϕ_A reaches saturation and the gate current, now designated by $i_{G,A,s}$ becomes limited by the nominal output resistance R_0 only, Fig. 2(A,3). The saturation component $i_{G,A,s}$ is usually several decades larger than the exciting component $i_{G,A,x}$ and therefore the saturation components $i_{G,A,s}$ and $i_{G,B,s}$ are shown Fig. 2(A,3) and 2(B,3), on different scales. For $\omega t>\alpha$ the gate ampere-turns of saturable reactor A swing way into the saturation branch 5, Fig. 2(A,2), of the B-H loop and return as the line voltage decreases. At $\omega t=\pi$, the gate ampere-turns $i_{G,A,s}N_G$ are again zero, and point 6 is reached.

In practice, the front of the gate current is not vertical, and the rising gate current produces by mutual induction a negative voltage in the control circuit. Then with the gate current declining, a positive voltage is induced in the control circuit. It is obvious that the rigorous tracking of the flux and control current may become quite complicated. For reasons of simplicity, it will be assumed that during the saturation interval the flux in the unsaturated core remains constant and the average value of the control current is zero. This condition is borne out by the oscillogram in Fig. 3; for different saturation angles, different core materials, and different core constructions, the agreement may not be as close. A further simplification is introduced by assuming the instantaneous values of the control current as zero during the saturation interval, Fig. 2(L,5).

At $\omega t=\pi$ the cores have traded their positions on the B-H loop with respect to $\omega t=0$; during the following half cycle, $\pi<\omega t<2\pi$, the events are analogous to those of the previous half cycle. The total gate ampere-turns are obtained by adding sections (A,3) to (B,3) and (A,4) to (B,4) in Fig. 2 after applying the proper scale factors. The total load current is obtained by adding all gate currents, Figs. 2(L,3), 2(L,4).

An oscillogram of load current i_L and control current i_C is shown in Fig. 3; the applied control voltage E_C is pure d-c. The test specimen consists of two spiral cores, having an outside diameter of $1^{25}/_{32}$ inches, and an inside diameter of $1^{15}/_{32}$ inches, a width of 1/2 inch, $e_{Fe}=$ 12.96 centimeters (cm), $A_{Fe}=0.457$ cm^2, a thickness of nickel alloy laminations of 0.002 inch, B-H loop at 60 cycles per second, according to Fig. 4; $N_G=1{,}000$ turns, $N_C=50$ turns; to eliminate reverse current, a tube rectifier *6AL5* is used. The excursion of the control current i_C is 20.7 milliamperes (ma) according to the oscillogram shown in Fig. 3, which value checks closely with equation 5. It will be noticed that the actual control current does not show as abrupt changes as the theory, Fig. 2(L,5), indicates, the mutual inductance between gate and control windings during the saturation interval; the mutual inductance is largely caused by the finite slope of the B-H loop in the saturation region.

Nominal Output Voltage E_0

The nominal output voltage e_0 is defined as the active voltage during the saturation interval

$$e_0=e_{1-N} \qquad e_0=e_{2-N} \tag{9}$$

$$\text{for } \alpha\leqslant\omega t\leqslant\pi \qquad \text{for } \alpha+\pi\leqslant\omega t\leqslant 2\pi$$

then the average nominal output voltage E_0 becomes, as shown in equations 9 and 1

$$E_0=\frac{1}{\pi}\int_\alpha^\pi \frac{\pi}{2}E\sin\omega t d(\omega t)=\frac{1+\cos\alpha}{2}E \tag{10}$$

where α is the angle at which saturation starts.

Load Current I_L

The load i_L consists of two parts: the saturation load current $i_{L,s}$ and the exciting load current $i_{L,x}$ as shown in Figs. 2 (L,3 and 4)

$$i_L=i_{L,s}+i_{L,x} \tag{11}$$

The saturation load current $i_{L,s}$ is

$$i_{L,s}=e_0/R_0 \tag{12}$$

where R_0 is called the nominal output resistance. The average value $I_{L,s}$ follows from equation 10

$$I_{L,s}=\frac{E}{R_0}\frac{1+\cos\alpha}{2} \tag{13}$$

The exciting load ampere-turns $I_{L,x}N_G$ are, as shown in Fig. 2(A,4), 2(B,4), and 2(L,4)

$$I_{L,x}N_G=2AT_c\frac{\alpha}{\pi} \tag{14}$$

From equations 11, 13, and 14 one finally obtains the average value of the load current I_L

$$I_L=I_{L,s}+I_{L,x}=\frac{E}{R_0}\frac{1+\cos\alpha}{2}+2AT_c\frac{\alpha}{\pi}\frac{1}{N_G} \tag{15}$$

Control Current

From Figs. 2(L,4 and 5), and equation 14, control ampere-turns I_CN_C are

$$I_CN_C=-\frac{1}{2}I_{L,x}N_G=-AT_c\frac{\alpha}{\pi} \tag{16}$$

hence

$$\alpha=-\frac{I_CN_C}{AT_c}\pi \tag{17}$$

Control Characteristic

By substituting equation 5 in 17, and equation 17 in 15, the load current as function of the control ampere-turns is obtained

$$\boxed{\begin{gathered}I_L=\underbrace{\frac{E}{R}\frac{1+\cos\frac{0.4I_CN_C}{H_cl_{Fe}}\pi}{2}}_{I_{L,s}}\underbrace{-2\frac{I_CN_C}{N_G}}_{I_{L,x}}\\ \text{for } -\frac{H_cl_{Fe}}{0.4\pi N_C}\leqslant I_C\leqslant 0\end{gathered}} \tag{18}$$

It can be recognized that the load current consists of a cosine function $I_{L,s}$ and a linear function $I_{L,x}$ of the control ampere-turns.

Except in the vicinity of $\alpha=\pi$, the predominant component is $I_{L,s}$. One concludes that the control characteristic will look very much like a cosine function and that the slope of the control characteristic is inversely proportional to coercive force H_c of the dynamic B-H loop.

Exciting Current and Lower End of Control Characteristic

For $\alpha=\pi$, $I_{L,s}=0$; the load current equals $I_{L,x}$ which, for $\alpha=\pi$, is called exciting current I_E, equations 18, 16, and 5

$$I_{L,x}=I_E=\frac{2AT_c}{N_G}=\frac{2H_cl_{Fe}}{0.4\pi N_G} \tag{19}$$

$$\text{for } \alpha=\pi$$

This is the same exciting current as is obtained when the reactor is connected as an ordinary, parallel-connected, saturable reactor.

When connected as amplistat, the con-

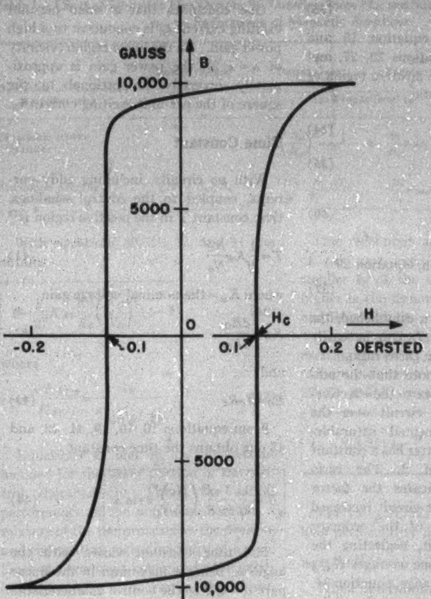


Fig. 4. Dynamic hysteresis loop, taken with sinusoidal voltage at 60 cycles per second

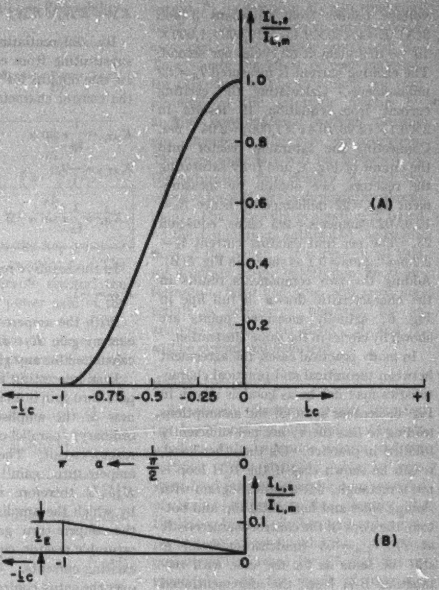


Fig. 5. Synthesis of control characteristic. (A) Per-unit saturation component of load current. (B) Per-unit exciting component of load current versus per unit control current

trol ampere-turns, which are necessary for the lower end of the control characteristic $\alpha=\pi$, $I_L=I_E$, point M in Fig. 5, follow from equations 16 and 5

$$\boxed{I_C N_C = -\frac{H_c l_{Fe}}{0.4\pi} \quad \text{for } \alpha=\pi} \tag{20}$$

These control ampere-turns can also be expressed by the exciting current, equation 16, $I_{L,x}=I_E$

$$\boxed{I_C N_C = -\frac{1}{2} I_E N_G \quad \text{for } \alpha=\pi} \tag{21}$$

Depending on whether H_c or I_E has been determined by direct measurement, equation 20 or 21 respectively is preferable.

Nondimensional (Normalized) Control Characteristic

When expressed in nondimensional form, the control characteristic of any amplistat with square-loop core material can be expressed, as far as the saturation component is concerned, by the same curve (cosine function). The only variation may occur because of the (linear) exciting component, the latter being usually very small.

Let us denote the per-unit control current of an amplistat by $\mathbf{i}_C$ and define

$$\mathbf{i}_C = \frac{I_C N_C}{H_c l_{Fe}/0.4\pi} \tag{22}$$

Further, let the per unit basis for the load and exciting currents be the load current $I_{L,m}$ obtained from equation 15 by letting $\alpha=0$

$$I_{L,m} = E/R_0 \tag{23}$$

then the control characteristic of the per unit saturation component $I_{L,s}/I_{L,m}$ becomes for any amplistat, Fig. 5(A)

$$\frac{I_{L,s}}{I_{L,m}} = \frac{1+\cos \mathbf{i}_C \pi}{2} \tag{24}$$

The control characteristic of the per unit exciting component $I_{L,x}$ is obtained by substituting equation 19 in 15; and equations 16, 22, and 5

$$\mathbf{i}_C = -\frac{\alpha}{\pi} \tag{25}$$

$$\frac{I_{L,x}}{I_{L,m}} = -\frac{I_E}{I_{L,m}} \mathbf{i}_C \tag{26}$$

Finally, denoting by $\mathbf{i}_E$, the per-unit exciting current

$$\mathbf{i}_E = \frac{I_E}{I_{L,m}} \tag{27}$$

one obtains for the per-unit exciting component $I_{L,x}/I_{L,m}$

$$\boxed{\frac{I_{L,x}}{I_{L,m}} = -\mathbf{i}_E \mathbf{i}_C} \tag{28}$$

as illustrated in Fig. 5(B). The total control characteristic is obtained by adding $I_{L,x}/I_{L,m}$ to $I_{L,s}/I_{L,m}$ as in Fig. 6.

With increasing size of the saturable reactor, the per-unit exciting current $\mathbf{i}_E$ declines. Amplistats of 100 watts output, 60 cycles per second, may be expected to have an $\mathbf{i}_E$ within 0.01 to 0.03.

Validity Check

To obtain a check for exciting current I_E, the previously mentioned reactor is connected as an ordinary, parallel-connected, saturable reactor. The applied voltage follows from equations 3 and 3(A) $E=4\times 60\times 0.457\times 10{,}000\times 1{,}000\times 10^{-8}=10.9$ volts at 60 cycles per second. The exciting current is measured $I_E=2.2$ milliamperes. Calculating the exciting current from equation 19 results in $2\times 0.1\times 12.96/(0.4\pi\times 1{,}000)=2.04$ ma. Connecting the saturable reactor into the circuit of Fig. 1, and fully saturating the reactors, one obtains by measurement $I_{L,m}=22$ milliamperes, hence $R_0=10.9/.022$ amperes $=495$ ohms, equation 23. The per unit exciting current $\mathbf{i}_E=2.2$ ma/22 ma $=0.1$ as shown in Fig. 5(B). Adding the two components results in the characteristic drawn in full line in Fig. 6; actually measured points are shown by circles in the same illustration.

In many practical cases, the agreement between theoretical and practical characteristics may not be as good as shown in Fig. 6, because some of the assumptions leading to this theory are not sufficiently fulfilled in practice. On the other hand, it can be shown that if the B-H loop is not a rectangle, but a parallelogram with sloping sides and horizontal top and bottom, the slope of the control characteristic at $I_{L,s}/I_{L,m}=0.5$ (midcharacteristic) is still the same as for the case with rectangular B-H loop; the aforementioned statement is restricted to slopes of the B-H loop larger than the square-loop permeability μ_{sq}, equation 52.

Negative Region

The control characteristic in the negative region is approximated without further proof by

$$I_L=\frac{1}{2}I_E-I_CN_C/N_G \tag{29}$$

$$\text{for}\quad\left(-I_{L,m}+\frac{H_c l_{Fe}}{0.4\pi N_G}\right)\frac{N_G}{N_C}\leqslant I_C\leqslant-\frac{H_c l_{Fe}}{0.4\pi N_C}$$

The normalized characteristic is the extension of the straight line, S (1,0) to M (-1, $\mathbf{i}_E$), as illustrated in Fig. 6.

Ampere-Turn Gain

By definition, the ampere-turn gain $K_{AT}=$

$$K_{AT}=\frac{dI_L}{dI_C}\frac{N_G}{N_C} \tag{30}$$

As in equation 15, K_{AT} can be split into saturation component $K_{AT,s}$ and exciting component $K_{AT,x}$

$$K_{AT,s}=\frac{dI_{L,s}}{dI_C}\frac{N_G}{N_C} \tag{31}$$

$$K_{AT,x}=\frac{dI_{L,x}}{dI_C}\frac{N_G}{N_C} \tag{32}$$

$$K_{AT}=K_{AT,s}+K_{AT,x} \tag{33}$$

By differentiating equation 15 and substituting from equations 23, 27, and 19, one obtains for the positive region of the control characteristic

$$K_{AT,s}=\frac{1}{\mathbf{i}_E}\pi\sin\alpha \tag{34}$$

$$K_{AT,x}=-2 \tag{35}$$

$$\boxed{K_{AT}=\frac{1}{\mathbf{i}_E}\pi\sin\alpha-2} \tag{36}$$

In the negative region, equation 29

$$K_{AT}=-1 \tag{37}$$

With the ampere-turn established, the current gain $K_I=dI_L/dI_C$ can be easily calculated for any given turns ratio.

It is interesting to note that the ampere-turn gain is an index to the effectiveness of the amplistat circuit over the ordinary, parallel-connected saturable-reactor circuit. The latter has a constant ampere-turn gain[11] of 2. The ratio $K_{AT,s}/2$ therefore indicates the factor by which the amplistat circuit increased the ampere-turn gain of the ordinary saturable-reactor circuit, neglecting the exciting currents. If one averages $K_{AT,s}$ over the entire control range, equation 34, this average $K_{AT,s\,avg}$ equals $2/\mathbf{i}_E$. On this basis, the amplistat increased the ampere-turn gain of the saturable reactor by the factor $1/\mathbf{i}_E$. This leads to the conclusion that the effectivness of the amplistat circuit is controlled by $1/\mathbf{i}_E$, and from this point of view a low per unit exciting current $\mathbf{i}_E$ is most desirable.

Power Gain

The power gain K_P is defined by

$$K_P=\left(\frac{dI_L}{dI_C}\right)^2\frac{R_L}{R_C}k_f^2 \tag{38}$$

Substituting from equation 30, the power gain can be expressed by the ampere-turn gain

$$K_P=K_{AT}^2\frac{R_L}{R_C}\left(\frac{N_C}{N_G}\right)^2k_f^2 \tag{39}$$

Substituting from equation 36 into 39, one obtains for the power gain

$$\boxed{K_P=\left(\frac{1}{\mathbf{i}_E}\pi\sin\alpha-2\right)^2\frac{R_L}{R_C}\left(\frac{N_C}{N_G}\right)^2k_f^2} \tag{40}$$

One concludes that a small per-unit exciting current $\mathbf{i}_E$ is conducive to a high power gain. In the linear region (vicinity of $\alpha=\pi/2$), the power gain is approximately inversely proportional to the square of the per-unit exciting current $\mathbf{i}_E$.

Time Constant

With no circuits, including eddy currents, coupled to the control windings, time constant T in the positive region is[12]

$$T=\frac{1}{2f}K_E\frac{N_C}{N_G} \tag{41}$$

where $K_E=$ the nominal voltage gain

$$K_E=\frac{dE_O}{dE_C} \tag{42}$$

and

$$E_C=I_CR_C \tag{43}$$

From equations 10, 16, 19, 41, 42, and 43 one obtains the time constant

$$\boxed{T=\frac{1}{2f}\frac{1}{I_E}\frac{E}{R_C}\left(\frac{N_C}{N_G}\right)^2\pi\sin\alpha} \tag{44}$$

The time constant varies with the angle α, being a maximum in the linear part $\alpha=\pi/2$ of the control characteristic. The time constant can be calculated from the exciting current I_E, which is, as previously mentioned, easily obtainable from measurement.

If a control characteristic is available, the time constant can be obtained from the ampere-turn gain.[13] First, the voltage gain and the saturation ampere-turn gain can be related by equations 10, 13, 31, 42, and 43

$$K_E=K_{AT,s}\frac{N_C}{N_G}\frac{R_O}{R_C} \tag{45}$$

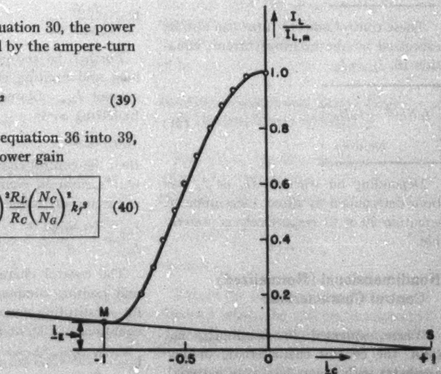


Fig. 6. Control characteristic for amplistat ($\mathbf{i}_E=0.1$). Solid curve from theory, Fig. 5; dots indicate measured points

Table I. Effect of Saturation Flux Density B_s, Coercive Force H_c, and Square-Loop Permeability μ_{sq} on Performance of Magnetic Amplifier

Operation of Magnetic Amplifier with	Power Output P_L	Power Gain K_P	Time Constant T	Dynamic Power Gain D_P
	are Proportional to			
Maximum power transfer	B_s^2	$\left(\frac{B_s}{H_c}\right)^2=\mu_{sq}^2.$	$\frac{B_s}{H_c}=\mu_{sq}$	$\frac{B_s}{H_c}=\mu_{sq}$
Maximum winding temperature	$(2B_s-B_{s,1})B_{s,}$	$\frac{(2B_s-B_{s,1})B_s,}{H_c^2}$	$\frac{B_s}{H_c}=\mu_{sq}$	$\frac{\left(2-\frac{B_{s,1}}{B_s}\right)B_{s,1}}{H_c}$

With equations 30, 31, 32, and 41 one obtains

$$\boxed{T=\frac{1}{2f}K_{AT}\frac{R_O}{R_C}\left(\frac{N_C}{N_G}\right)^2(1+\kappa)} \tag{46}$$

where

$$\kappa=\frac{-K_{AT,x}}{K_{AT}}=\frac{\mathfrak{i}_E}{\frac{\pi}{2}\sin\alpha-\mathfrak{i}_E} \tag{47}$$

Equations 44 and 46 are not recommended for the lower portion of the control characteristic. There, the actual performance of the amplistat becomes a mixture of the performance in the positive and the exciting region (vicinity of $\alpha=\pi$). Since the transient response in the exciting region may vary over a wide range, depending on the per unit excitation and the magnitude of the control voltage, the time constant of the magnetic amplifier cannot be dependably described by the method given when approaching or operating in the exciting region.

Dynamic Power Gain

The dynamic power gain D_P is defined by[14]

$$D_P=K_P/T \tag{48}$$

From equations 39 and 46 is obtained

$$\boxed{D_P=2fK_{AT}\frac{R_L}{R_O}\frac{1}{1+\kappa}k_f^2} \tag{49}$$

The ratio R_L/R_O is subject only to minor variations, namely, from 0.5 for small reactors to 0.8 to 0.9 for large reactors. The term $1+\kappa$ usually is near unity in the linear region. Hence, the only dominating terms are frequency f and ampere-turn gain K_{AT}.

At a given frequency, the ampere-turn gain is a basic parameter indicating the dynamic power gain.

Substituting into equation 49 from 36, and 47

$$\boxed{D_P=2f\frac{1}{\mathfrak{i}_E}\frac{R_L}{R_O}\frac{1}{(1+\kappa)^2}\pi\sin\alpha k_f^2} \tag{50}$$

One recognizes again the key position of per unit exciting current $\mathfrak{i}_E$. The smaller $\mathfrak{i}_E$ of the saturable reactor, the higher is the dynamic power gain of the amplistat circuit.

In view of the importance of per unit exciting current $\mathfrak{i}_E$, it is expressed in terms of core parameters. Utilizing equations 27, 23, 3, 3(A), and 19

$$\boxed{\mathfrak{i}_E=R_O\left(2f0.4\pi\frac{B_s}{H_c}\frac{A_{Fe}}{l_{Fe}}N_G^2 10^{-8}\right)^{-1}} \tag{51}$$

The ratio B_s/H_c is called the square-loop permeability μ_{sq}

$$\mu_{sq}=B_s/H_c \tag{52}$$

When determining μ_{sq} from a real B-H loop, the term B_s in equation 52 is replaced by the amplitude of the actual flux density.

The square-loop permeability is an abstract quantity; it shall not be construed that the performance of a square-loop core material can be approximated by a single-valued, straight B-H line with the slope of μ_{sq}. The practical significance of μ_{sq} can be observed from Table I.

Evaluation of Core Properties

The most important properties of magnetic amplifiers usually are: power output P_L, power gain K_p and time constant T, the latter two quantities being expressed by the dynamic power gain D_P. To evaluate the effect of core material, the procedure will be to replace one core material in the saturable reactor by another core material without causing any change in the core configuration. The core materials will be described by the saturation flux density B_s and the coercive force H_c of the dynamic hysteresis loop. The following evaluation is based on the assumptions outlined at the beginning of the paper. It is recognized that for accurate design calculations more information regarding core and core materials is necessary; however, for an analysis of first-order effects, B_s and H_c will be considered sufficient.

Power Output P_L

To obtain maximum power transfer, the load must be matched to the gate winding of the saturable reactor and the rectifier. With matched, that is, constant load resistance, and a core material of higher saturation flux density B_s, the line voltage E can be increased in proportion, equations 3 and 3(A); the power output will increase by E^2 and B_s^2 respectively. As the result of increased load current, the winding temperature will increase and for a certain voltage, called E_1, corresponding to $B_{s,1}$, the maximum permissible temperature of the winding will be reached. If one wishes to replace the core material by another one such that $B_s>B_{s,1}$, the line voltage can again be increased; however, not to ex-

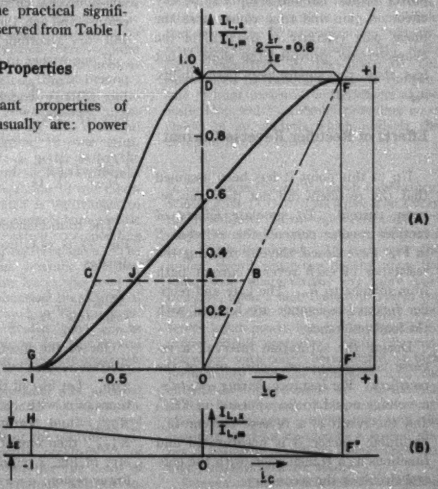


Fig. 7. Synthesis of the control characteristic in the presence of rectifier reverse current

ceed the maximum winding temperatures, the load current must be kept constant. This means that the load resistance R_L must be increased. It can be shown that the load resistance must increase according to $R_L/R_0=1-(1/2)E_1/E$. Then, the power output increases by $(2E-E_1)E_1$, and the corresponding flux density relations are shown in Table I.

POWER GAIN K_P

From equation 38 it is deduced that K_P is proportional to P_L and inversely proportional to I_C^2, and hence inversely proportional to H_c^2, equations 16 and 5.

TIME CONSTANT T

The time constant can be written in the following form, equations 44, 3, 3(A), and 19

$$T=0.4\pi\mu_{sq}\frac{A_{Fe}}{l_{Fe}}N_C^2 10^{-8}\frac{1}{R_C}\pi\sin\alpha \tag{53}$$

The time constant is proportional to the square-loop permeability μ_{sq}.

DYNAMIC POWER GAIN D_P

Dynamic power gain follows from equation 48; see Table I.

One of the conclusions is that a small coercive force H_c of the dynamic loop is always desirable, and that a high saturation flux density B_s is desirable in most cases.

An extension of this theory to the B-H loop with sloping sides but horizontal top and bottom indicates that a finite slope has no effect on the maximum power output, nor has a finite slope any effect on gain and time constant in the linear part (vicinity of $\alpha=\pi/2$) of the characteristic, provided the slope is not less than the square-loop permeability μ_{sq}.

Effects of Rectifier Reverse Current

Up to this point it has been assumed that the rectifiers did not pass any reverse current. To simulate effects of rectifier reverse current, the switches S in Fig. 1 are closed, thereby reducing the resistance of each reverse current path from infinity to R_r. The gate and rectifier forward resistance are lumped with the load resistance.

During the saturation interval, a reverse voltage appears across one of the rectifiers. For instance, during $\alpha<\omega t<\pi$, a voltage equal to $2e_0$ appears on REC_2 and the result is a reverse current $i_{G,r}$. Saturable reactor B is unsaturated and functions as a transformer, with the control circuit as the secondary

$$-i_{G,r}N_G=i_{C,r}N_C \tag{54}$$

Since the control-circuit resistance R_C is small, the voltage drop on R_C is small, and hence the rate of flux change $d\phi_B/dt$ is also small. As a result, the eddy currents will be comparatively small, and hence the B-H loop for the transformation of reverse current will, in general, be much narrower than shown in Fig. 4; for reasons of simplification, its width will be assumed zero for the transformation of reverse current.

As shown in the Appendix, the per unit control ampere-turns can be expressed by

$$\boxed{\mathbf{i}_C=-\frac{1}{\pi}\arccos\left(2\frac{I_{L,s}}{I_{L,m}}-1\right)+2\frac{\mathbf{i}_r}{\mathbf{i}_E}\frac{I_{L,s}}{I_{L,m}}} \tag{65}$$

with $\mathbf{i}_r$ being the per unit reverse current, equation 60. This equation permits the rapid construction of the control characteristic. The first addend of equation 65 (Fig. 7, GCD) is identical with the cosine function of Fig. 5(A); the second addend represents a straight line OF (Fig. 7), such that $DF=2\mathbf{i}_r/\mathbf{i}_E$. By adding the abcissas (by making $CJ=AB$), one obtains quickly the characteristic GJF, representing the saturation component of the load current in the presence of reverse current.

One recognizes that the reverse current causes a reduction of gain[15]; furthermore, reverse current makes the characteristic cross the ordinate axis. The higher the reverse current, the lower is the crossing point. If rectifiers are used whose reverse current varies with temperature and age, the control characteristic will vary accordingly.

If one wishes to add the exciting component $I_{L,x}/I_{L,m}$, the characteristic of Fig. 5(B) is stretched to cover the new base GF', Fig. 7

$$\boxed{\frac{I_{L,x}}{I_{L,m}}=\mathbf{i}_E\left(2\frac{\mathbf{i}_r}{\mathbf{i}_E}-\mathbf{i}_C\right)\Big/\left(2\frac{\mathbf{i}_r}{\mathbf{i}_E}+1\right)} \tag{66}$$

The total characteristic (not shown in Fig. 7) is obtained by adding the ordinates

$$\frac{I_L}{I_{L,m}}=\frac{I_{L,s}}{I_{L,m}}+\frac{I_{L,x}}{I_{L,m}} \tag{67}$$

One of the most important effects of rectifier reverse current is a reduction of gain. Let us call the saturation ampere-turn gain with rectifier reverse current $K_{AT,s,r}$ and without reverse current $K_{AT,s,o}$, then one obtains from the geometry of Fig. 7, equation 34, and for the linear region, $\alpha=\pi/2$

$$\boxed{\frac{K_{AT,s,r}}{K_{AT,s,o}}=\frac{1}{1+\pi\dfrac{\mathbf{i}_r}{\mathbf{i}_E}}} \tag{68}$$

It is important to recognize that the gain reduction is caused by the ratio $\mathbf{i}_r/\mathbf{i}_E$. When cores of higher performance (smaller $\mathbf{i}_E$) are used, rectifiers of lower reverse current (smaller $\mathbf{i}_r$) must be applied if one does not wish a disproportionate gain reduction. Furthermore, since rectifier reverse current is usually proportional to the area of the rectifier, the application of a rectifier which is rated for a current in excess of the actual current produces more reduction of gain.

The reduction ratio of equation 68 applies as a good approximation not only to the ampere-turn gain, but also to the time constant, equation 44, $\alpha=\pi/2$, and the dynamic power gain, equation 50, $\alpha=\pi/2$. To obtain the power gain, equation 40, the square of the reduction ratio is applied. When determining power gain and dynamic gain from equations 39 and 49, the reduction ratio does not enter these equations.

The ampere-turn gain $K_{AT,s,r}$ in the presence of reverse current can be calculated for the linear part of the characteristic from equations 68 and 34

$$\boxed{K_{AT,s,r}=\frac{1}{\dfrac{\mathbf{i}_E}{\pi}+\mathbf{i}_r}} \tag{69}$$

It is recognized that core and rectifier quality jointly determine the performance of the amplistat. There is little to be gained by using a high-performance core (small $\mathbf{i}_E$) together with low-quality rectifiers (large $\mathbf{i}_r$). To realize the full advantage of high-performance cores, such cores must be matched to rectifiers of correspondingly small reverse current.

Appendix

The reverse gate current $i_{G,r}$ can be approximated by the reverse voltage e_r

$$i_{G,r}=e_r/R_r \tag{55}$$

The reverse voltage =

$$e_r=-2e_0 \tag{56}$$

and the average reverse gate current

$$I_{G,r}=\frac{1}{\pi}\int^{\pi}i_{G,r}d(\omega t)=\frac{1}{\pi}\frac{1}{R_r}\times\int_{\alpha}^{\pi}e_r d(\omega t)=-2\frac{E_0}{R_r} \tag{57}$$

From equations 54 and 57

$$2\frac{E_0}{R_r}N_G = I_{C,r}N_C \tag{58}$$

The maximum, average reverse current $I_{r,m}=$

$$I_{r,m} = -2E/R_r \tag{59}$$

Let us denote by $\mathfrak{i}_r$ the per-unit reverse current

$$\frac{|I_{r,m}|}{I_{L,m}} = \mathfrak{i} \tag{60}$$

From equations 23 and 59

$$R_0/R_r = \frac{1}{2}\mathfrak{i}_r \tag{61}$$

Substituting into equation 58 for $E_0 = I_{L,s}R_O$, equation 12, and from equation 61, one obtains the control ampere turns caused by reverse current

$$I_{C,r}N_C = \mathfrak{i}_r I_{L,s}N_G \tag{62}$$

The control ampere-turns in the absence of reverse current are (equations 13, 23, 17, 16, and 19)

$$I_C N_C = \frac{H_d l_{Fe}}{0.4\pi}\left[-\frac{1}{\pi}\arccos\left(2\frac{I_{L,s}}{I_{L,m}} - 1\right)\right] \tag{63}$$

In the presence of reverse current, the total control ampere-turns are obtained as the sum of the control ampere-turns, equations 62 and 63. In view of equations 19 and 27 one obtains for the total control ampere-turns

$$I_C N_C = \frac{H_d l_{Fe}}{0.4\pi}\left[-\frac{1}{\pi}\arccos\left(2\frac{I_{L,s}}{I_{L,m}} - 1\right) + 2\frac{\mathfrak{i}_r}{\mathfrak{i}_E}\frac{I_{L,s}}{I_{L,m}}\right] \tag{64}$$

By substituting equation 22 into 64 one obtains

$$\mathfrak{i}_C = -\frac{1}{\pi}\arccos\left(2\frac{I_{L,s}}{I_{L,m}} - 1\right) + 2\frac{\mathfrak{i}_r}{\mathfrak{i}_E}\frac{I_{L,s}}{I_{L,m}} \tag{65}$$

The per-unit exciting component is obtained by spreading the straight-line characteristic of Fig. 5(B) below the new characteristic $-1 < \mathfrak{i}_C < 2\mathfrak{i}_r/\mathfrak{i}_E$

$$\frac{I_{L,x}}{I_{L,m}} = \mathfrak{i}_E\left(2\frac{\mathfrak{i}_r}{\mathfrak{i}_E} - \mathfrak{i}_C\right)\Big/\left(2\frac{\mathfrak{i}_r}{\mathfrak{i}_E} + 1\right) \tag{66}$$

The total per-unit load current becomes the sum of $I_{L,s}/I_{L,m}$ and $I_{L,x}/I_{L,m}$, equation 11. In the last derivation it has been tacitly assumed that the reverse current is so small with respect to the load current that its subtractive effect on the latter is negligible.

References

1. A New Theory of the Magnetic Amplifier, **A. G. Milnes.** *Proceedings*, Institution of Electrical Engineers, London, England, vol. 97, pt. II, Aug. 1950, pp. 460–63.

2. The Amplistat and its Application, Part I, **H. M. Ogle.** *General Electric Review*, Schenectady, N. Y. vol. 53, Feb. 1950, Fig. 1, pp. 32–35.

3. Standardized Magamps, **R. W. Roberts.** *Westinghouse Engineer*, Pittsburgh, Pa., Nov. 1952, p. 195.

4. The Transductor and its Applications, **U. Lamm.** *Journal* Almanna Svenska Electriska Aktiebologet, London, England, 1939, pp. 66–80.

5. Self-saturation in Magnetic Amplifiers, **W. J. Dornhoefer.** *AIEE Transactions*, vol. 68, pt. II, 1949, pp. 835–50.

6. A Cathode-Ray B-H Tracer, **Joseph Zamsky.** *Electrical Engineering*, vol. 66, July 1947, pp. 678–80.

7. Die Messung der magnetischen Eigenschaften, **F. Koppelmann.** *A.E.G. Mitteilungen*, Berlin, Germany, vol. 42, 7/8, 1952, p. 172.

8. Dynamic Hysteresis Loop Measuring Equipment, **H. W. Lord.** *AIEE Transactions*, vol. 72, pt. I, Sept., 1952, pp. 269–72.

9. Evaluation of Core Materials for Magnetic Amplifiers—I, **D. C. Dieterle.** *Electrical Manufacturing*, New York, N. Y., Jan. 1953, pp. 68–73, 274, 276.

10. Dynamic Hysteresis Loops of Several Core Materials Employed in Magnetic Amplifiers, **H. W. Lord.** *AIEE Transactions*, vol. 72, pt. I, March 1953, pp. 85–88.

11. Transient Response of Saturable Reactors with Resistive Load, **H. F. Storm.** *AIEE Transactions*, vol. 70, pt. I, 1951, pp. 95–102.

12. The Transductor Amplifier, **U. Krabbe.** *Lindhska Boktryckeriet*, Orebro, Sweden, 1947, p. 87.

13. The Figure of Merit of Magnetic Amplifiers, **J. T. Carleton, W. F. Horton.** *AIEE Transactions*, vol. 71, pt. I, 1952, pp. 239–45.

14. L'Amplidyne et ses applications, **M. A. Valentin.** *Revue d'Electricité et de Mécanique*, Paris, France, no. 74, July—Sept. 1948, pp 1–16.

15. The Effective Feedback Ratio of Magnetic Amplifiers, **L. A. Finzi, G. F. Pittman, Jr., H. L. Durand.** *AIEE Transactions*, vol. 71, pt. I, April 1952, pp. 157–64.

◆

Discussion

Harold W. Lord (General Electric Research Laboratory, Schenectady, N. Y.): The analysis of the magnetic amplifier which Dr. Storm has presented is certainly a step forward in this field. The proof of its value lies in the very close agreement he shows in Fig. 6 between theory and measurement. I understand that Dr. Storm has obtained similar results for cores of several other materials.

I would like to show that this analysis has some limitations which, incidentally, Dr. Storm freely admits in his paper. It assumes the control current to be zero during the conduction period. A very rectangular hysteresis loop material which nearly conforms to this assumption is 65 Permalloy. A pair of 0.005-inch-thick strip-wound core specimens of such a material were used for checking the validity of Dr. Storm's analysis. They had static loop coercive forces of approximately 0.02 oersted, and 60-cycle dynamic loop coercive forces of approximately 0.17 oersted. In Fig. 8 the circles indicate experimentally determined test points. Curve *A* was calculated using equation 18. Curve *B* was calculated using a modification of his analysis, which is equivalent to shearing the top of the control characteristic curve to the left to an extent corresponding to the static coercive force, and using the difference between the dynamic and static coercive forces as the value for H_c in equation 18. Note that this minor modification improves the accuracy of the results for this type of material.

The use of this modification for materials having very rectangular static hysteresis loops of the order of 50 per cent of the dynamic loop widths were not so successful. Fig. 9 shows the results for 0.002-inch-thick strip-wound Deltamax cores having static coercive forces of approximately 0.09 oersted and 60-cycle dynamic coercive forces of approximately 0.17 oersted. The nomenclature is similar to that for Fig. 8.

The measured control curve for this material falls between the two calculated curves *A* and *B*. After observing this, a second modification of equation 18 was tried, which empirically gives a better fit for the very rectangular hysteresis loop materials. For this calculation, equation 18 was modified in a manner which is equivalent to shearing the top of the control characteristic curve to the left to an extent corresponding to one-half the static coercive force. The difference between the dynamic and one-half the static coercive forces was used as the value for H_c in equation 18. Curve *C* of Fig. 9 was calculated in this manner. This modification provided good checks with experimental curves for the several other very rectangular loop core materials tested.

This empirical approach was adopted after studying the manner in which the wave form of the current in the control circuit varied throughout the control range. Fig. 10 shows composite oscillograms of such current wave forms for the cores of 65 Permalloy and Deltamax referred to. This demonstrates that an analysis which would account for the indicated changes in the wave form of the control current probably would be very complex and would involve, in addition to the static and dynamic hysteresis loops, such parameters as saturation-region permeabilities, initial permeabilities, and load currents. It is probable that one of the latter parameters, namely, saturation-region permeability, had sufficient magnitudes in the materials Dr. Storm used for validity checks to compensate in large measure for the effects of the static loop widths, thereby providing good checks of an analysis which neglects the effects of both of them.

D. G. Scorgie (Naval Research Laboratory, Washington, D. C.): This paper is on a very timely subject. In several recent papers[1–3] magnetic circuits with unique properties have been developed. In each case the unique circuit properties depend upon the square-loop characteristic or, more particularly, upon a high remanent flux in the magnetic cores. High permeability conserves input power but is not essential to proper circuit operation. Quite the contrary, for, with cores characterized by high permeability and low remanent flux, the circuits do not operate at all. As the title of Dr. Storm's paper indicates, he recognizes the disparity which exists, and the resulting need for developing a special theory of magnetic amplifiers where cores with high remanent flux are used.

Our remarks are concerned with additions

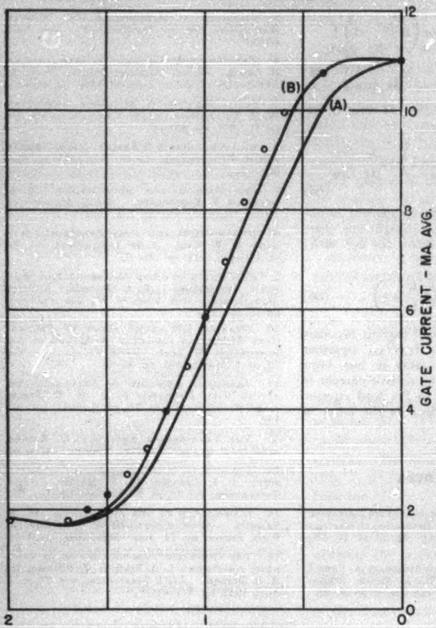


Fig. 8. Magnetic amplifier control characteristic curves for 65 Permalloy cores

Curve A—calculated from H. F. Storm's analysis
Curve B—calculated from first modification
Circles show test points

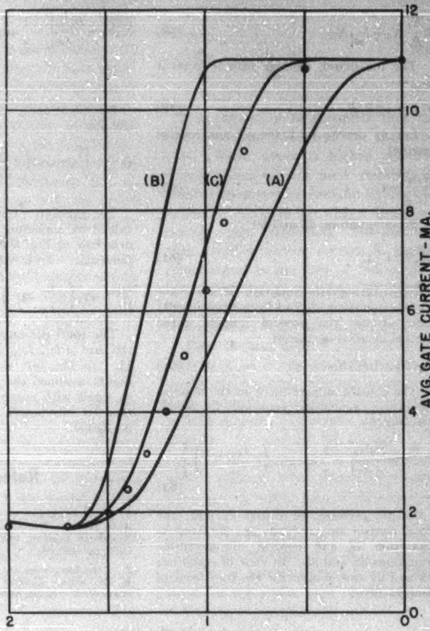


Fig. 9. Magnetic amplifier control characteristic curves for Deltamax cores

Curve A—calculated from H. F. Storm's analysis
Curve B—calculated from first modification
Curve C—calculated from second modification
Circles show test points

to, rather than changes in, the substance of what Dr. Storm has already said. The heart of his analysis lies in the two sections which discuss the exciting interval and the saturation interval. The control circuit resistance and control voltage were assumed negligible and the analysis showed that each half cycle the cores trade their positions on the B-H loop, thus describing a steady-state condition. It should be pointed out that under the basic assumptions the steady-state output current depends only upon the initial condition of saturation of the two cores. That is to say, if points 1 and 2, Fig. 2 (A, 2), were chosen at a different arbitrary flux level, the analysis would have shown a correspondingly different steady-state output current. As shown in Fig. 2, the average control current is determined by the point at which the saturation of core *B* occurs, hence is also determined by the initial conditions of core saturation. The relationship between output and control current given by equation 18 is valid for the assumed conditions. An interesting result is found when to the conditions assumed by Dr. Storm one adds current flow in a separate d-c bias winding. Instead of simply shifting the control characteristic bodily to the right, the bias current increases its slope. The magnitudes of both the exciting component and the saturation component, Fig. 5, are decreased at the lower end of the characteristic, leaving the upper end unaffected.

If $E_c=0$ and a small control circuit resistance R_c is introduced, control current flow through R_c causes a voltage drop in such a direction as to drive the amplifier to full saturation output. A negative E_c must be added to shut the amplifier off. Yet an analysis which takes into account E_c and

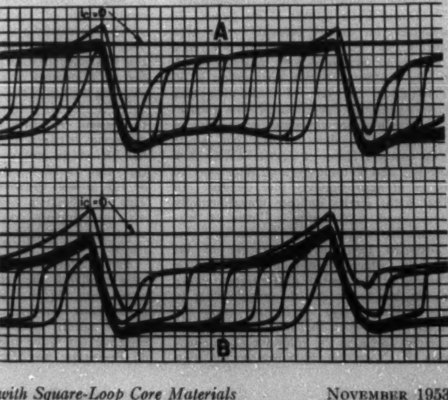


Fig. 10. Composite oscillograms of control circuit current

A—For 65 Permalloy wound cores of 0.005-inch × 0.5-inch strip
B—For Deltamax wound cores of 0.002-inch × 0.5-inch strip

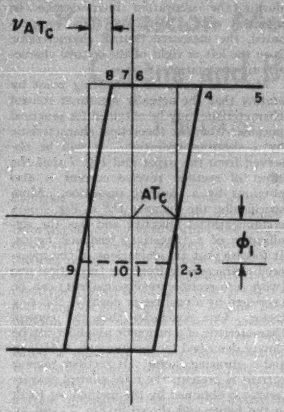


Fig. 11. Parallelogrammatic B-H loop

R_c as small but finite quantities reveals some strange behaviour for amplifiers with ideal rectifiers and perfectly square-loop cores. In fact one is forced to conclude that these idealized amplifiers have unstable regions in their voltage transfer characteristics. The substantiating argument is too lengthy to be given here but is included in a paper which has been submitted to the AIEE for publication in the *Transactions*. When one assumes a small fixed control resistance R_c and dries to derive a transfer characteristic between output and control voltages, the assumption of square-loop cores and perfect rectifiers leads to stability difficulty. This is not the triggering effect commonly observed in the lower portion of the transfer characteristic, but rather is a theoretically infinite voltage gain calculated over the entire range of outputs.

In practice there are other effects, mentioned by Storm in his section on rectifier leakage, which would prevent this infinite voltage gain even with square-loop cores, namely rectifier leakage and the narrowing of the B-H loop as the rate of change of flux decreases. When E_c is changed so as to increase the amplifier output, the B-H loops become slightly narrower and more back voltage is applied to the rectifier in series with the resetting core. The resulting change in the I_cR_c voltage drop is an effective negative voltage feedback which tends to reduce the voltage gain and prevent this type of instability.

In setting up the physical assumptions, E_c and R_c were assumed equal to zero while the current equations were derived. Later the derived control current was multiplied by R_c to determine E_c. This line of reasoning overlooks effects which can become important, especially in very high gain amplifiers. A more careful consideration of E_c and R_c would form a valuable addition to the literature.

REFERENCES

1. ON THE MECHANICS OF MAGNETIC AMPLIFIER OPERATION, Robert A. Ramey. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1214–23.

2. ON THE CONTROL OF MAGNETIC AMPLIFIERS, R. A. Ramey. *Ibid.*, pp. 2124–28.

3. LEAD NETWORKS UTILIZING A SATURABLE CORE MEMORY, D. G. Scorgie. *AIEE Transactions*, vol. 71, pt. II, Nov. 1952, pp. 297–302.

R. A. Phillips (General Electric Company, Schenectady, N. Y.): With his clear-cut analytical approach, the author should dispel many false notions which have appeared in the literature in the past about the influence of the dynamic loop upon the control characteristic of a magnetic amplifier.

Dr. Storm mentions that, if the B-H loop is not a rectangle but a parallelogram with sloping sides and horizontal top and bottom, the slope of the saturation component of the control characteristic at $I_{L,s}/I_{L,m}=0.5$ (mid-characteristic) is the same as for the case with rectangular B-H loop. I feel that a short proof would be of help to others in the field. The complete control characteristic is also developed.

In the interest of brevity, I will use the author's equations directly where applicable. Actually it is necessary to write but three new equations and draw a B-H loop, shown in Fig. 11, where a new constant ν is defined. The exciting components of the load current in cores A and B are

$$i_{L,xA}=\frac{AT_C}{NG}\left[1+\frac{\phi_A}{\phi_s}\right]0\leqq\omega t\leqq\alpha \tag{70}$$

$$i_{L,xB}=\frac{AT_C}{N_G}\left[1-\frac{\phi_B}{\phi_s}\right]0\leqq\omega t\leqq\alpha \tag{71}$$

$$i_{L,x}=i_{L,xA}+i_{L,xB} \tag{72}$$

Substituting equations 70, 71, 7, and 8 into 72

$$i_{L,x}=\frac{AT_C}{N_G}\left[2+\left(\frac{\phi_1}{\phi_s}+1-2\cos\omega t\right)\right] \tag{73}$$

The average exciting component of load current is

$$I_{L,x}=\frac{1}{\pi}\int_0^\alpha i_{L,x}d(\omega t) \tag{74}$$

Substituting equation 73 into 74 and integrating

$$I_{L,x}=\frac{AT_C}{N_G}\left\{\left[(2+\nu)+\frac{\phi_1}{\phi_s}\right]\alpha-2\nu\sin\alpha\right\} \tag{75}$$

The quantity ϕ_1/ϕ_s may be determined by evaluating equation 8 at $\omega t=\alpha$. From this

$$\frac{\phi_1}{\phi_s}=\cos\alpha \tag{76}$$

Substituting equation 76 into 77

$$I_{L,x}=\frac{AT_C}{\pi N_G}(2+\nu)\alpha+\nu\alpha\cos\alpha-2\nu\sin\alpha \tag{77}$$

The saturation component is identical to equation 13

$$I_{L,s}=\frac{E}{2R_0}(1+\cos\alpha) \tag{78}$$

The control current which is flowing during the portion of the cycle for $0\leqq\omega t\leqq\alpha$ is

$$i_CN_C=-N_Gi_{L,xB} \tag{79}$$

Substituting equations 71 and 8 into 78

$$i_C=-\frac{AT_C}{N_C}1-\nu\cos\omega t \tag{80}$$

The average control current may be determined from an equation similar to 68

$$I_C=-\frac{AT_C}{\pi N_C}[\alpha-\sin\alpha] \tag{81}$$

Going now to per-unit quantities used by the author

$$\frac{I_{L,x}}{I_{L,m}}=\frac{\mathbf{i}_E}{2\pi}[(2+\nu)\alpha+\nu\alpha\cos\alpha-2\nu\sin\alpha] \tag{82}$$

$$\frac{I_{L,s}}{I_{L,m}}=\frac{1+\cos\alpha}{2} \tag{83}$$

$$\mathbf{i}_C=-\frac{1}{\pi}[\alpha-\nu\sin\alpha] \tag{84}$$

Equations 82, 83, and 84 readily reduce to those shown in the paper for $\nu=0$. Fig. 12 shows the results for $\nu=0.5$ as well for a comparison.

Ampere-turn gain may be readily determined by differentiating equations 77, 78, and 81 and substituting them into equations 31 and 32.

$$K_{AT,s}=\frac{1}{\mathbf{i}_E}\pi\left[\frac{\sin\alpha}{1-\nu\cos\alpha}\right] \tag{85}$$

$$K_{AT,x}=-\left[\frac{(2-\nu)-\nu\cos\alpha-\alpha\nu\sin\alpha}{1-\nu\cos\alpha}\right] \tag{86}$$

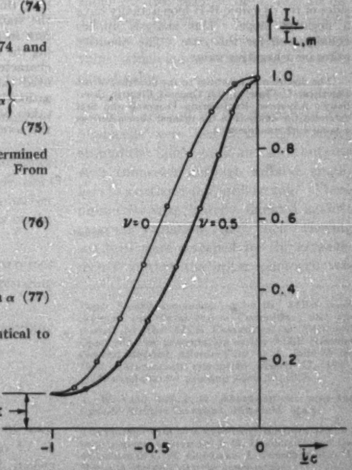


Fig. 12. Control characteristic for amplistat for $\nu=0$, 0.5

Table II

v	$K_{AT\cdot s}$	$K_{AT\cdot x}$
0	$\frac{\pi}{\mathfrak{i}_E}$	-2
0.5	$\frac{\pi}{\mathfrak{i}_E}$	-1.72
1.0	$\frac{\pi}{\mathfrak{i}_E}$	-1.43

At mid-characteristic ($\alpha=\pi/2$), equations 85 and 86 reduce to

$$K_{AT,s}\Big|_{\alpha=\frac{\pi}{2}}=\frac{\pi}{\mathfrak{i}_E} \qquad (87)$$

$$K_{AT,x}\Big|_{\alpha=\frac{\pi}{2}}=-(2-v)+\frac{\pi}{2} \qquad (88)$$

A tabular evaluation of equations 87 and 88 is shown in Table II. In a practical case $\mathfrak{i}_E$ lies between 0.01 and 0.1.

Several conclusions may now be drawn. Table II gives numerical values for the saturation component and the exciting component of ampere-turn gain. Since in the practical case exciting current is small, the contribution of $K_{AT,x}$ (the exciting component) to the total ampere-turn gain is decidedly a second-order effect. One may then say that the saturation component of the ampere-turn gain of a magnetic amplifier is not affected and the total ampere-turn gain is affected hardly at all by the deviation from the vertical of the steep sides of the B-H loop as long as the differential permeability of the steep sides of the B-H loop exceeds the square loop permeability, $\mu_{sq}(B_s/H_c)$. In a feedback control system the important thing about the shape of the control characteristic of a component magnetic amplifier is its slope over its operating region and not its position. Fig. 11 shows that the slope over the operating region (mid-characteristic) is changed very little by the sloping sides of the B-H loop.

In some magnetic materials the steep sides of the dynamic B-H loop actually have a negative slope. This analysis applies equally well for this case. The quantity takes on a negative value.

(The discusser is indebted to the students of the Electrical C Class of the General Electric Company's Advanced Engineering Program who first extended the analysis in the manner shown here as a homework assignment.)

H. F. Storm: I wish to thank R. A. Phillips for his thorough proof demonstrating that the saturation component of the gain of a magnetic amplifier at mid-characteristic is invariant with respect to the slope of the dynamic B-H loop, provided the slope is larger than the square loop permeability.

I wish to express gratitude to H. W. Lord for his efforts to make validity checks of the theory, and to find empirical correction factors for the theory. I believe that the concluding remark of Mr. Lord's discussion would have been more revealing by stating that the theory is based on the case where the opposing effects of static loop and mutual inductance are cancelling each other, as will become clearer from the following:

Let us assume zero gate reactance during the saturation interval. The cores will operate during this interval on a B-H loop which is much narrower than the dynamic B-H loop, and which, as Mr. Lord suggests, is the static B-H loop. If one denotes the ratio of widths between static and dynamic B-H loops by $\mathfrak{h}$, equation 18 becomes

$$I_L=\frac{E}{R_0}\frac{1}{2}\left(1+\cos\frac{\frac{0.4\pi I_C N_C}{H_c l_{Fe}}+\mathfrak{P}}{1-\mathfrak{P}}\right)-I_E\,\frac{\frac{0.4\pi I_c N_c}{H_c l_{Fe}}=\mathfrak{P}}{1-\mathfrak{P}} \qquad (89)$$

Equation 89 describes a cosine function of shorter period, but having a common end point and $\mathfrak{P}_C=-1$; see Fig. 5(A). One concludes that the introduction of the static B-H loop shifts the control characteristic to the left.

In the presence of mutual inductance during the saturation interval, voltages are induced from the gate winding into the control winding. The interesting part of these voltages is the positive one. It will be seen from Fig. 3 and Fig. 9(A) and (B) that this voltage counteracts the effect of the static loop by driving the control current in a positive direction, thereby reducing the deviations of the theoretical control characteristic (equation 18) from the measured values. In many cases, full compensation between these two opposing effects takes place, as assumed in the paper, and the average value of the control current during the saturation interval can be assumed zero. When not fully compensated, the measured control characteristic is to the left or right of the control characteristic of equation 18.

Mr. Lord made an interesting point by stating that the actually measured control characteristic may be obtained for practical purposes from the theoretical characteristic by a shearing operation. It will be observed from the paper and Fig. 7 that the effect of rectifier reverse current is also obtained by a shearing operation. More specifically, the distance DF in Fig. 7 is a nondimensional quantity and has the significance of a (negative) feedback factor. It follows that the deviations of the measured characteristic from the theoretical one (with zero rectifier reverse current) can be expressed by a positive or negative shearing factor. One concludes that the control characteristic of a magnetic amplifier can be closely described by the dynamic B-H loop and a shearing factor. If rectifier reverse current is present, the total control characteristic is obtained by the method set forth in the paper and shown in Fig. 7, and by using a distance DF comprising both the rectifier reverse current factor and the shearing factor.

The author does not care to take issue with the various statements made by Mr. Scorgie, except his concluding paragraph. As pointed out in the synopsis of the paper, it applies to the so-called low-control circuit-impedance type. This type is characterized by a control circuit resistance which is small by comparison with the unsaturated control winding reactance. A typical value of the ratio of control circuit resistance to control winding reactance is in the vicinity of 0.01. During the exciting interval, one core is excited by the other. Since the control circuit resistance is only approximately 0.01 of the exciting reactance of the control winding, the control circuit resistance has only a minor effect on the excitation of the second core. Once the excitation has been thus determined, the control voltage E_C and the control circuit resistance R_C can be related, as shown by equation 43.

Without giving any substantiation, Mr. Scorgie claims that "This line of reasoning overlooks effects which can become important, especially in high gain amplifiers." If this is so, the author can only regret that Mr. Scorgie did not substantiate this claim.

Propagation Mechanism of Impulse Corona and Breakdown in Oil

T. W. LIAO
MEMBER AIEE

J. G. ANDERSON
ASSOCIATE MEMBER AIEE

Synopsis: The breakdown mechanism of oil dielectric has not been fully understood. Information on impulse corona and breakdown are particularly meager. Breakdown theories postulated in the past are, in general, based on voltage measurements and are mostly restricted to uniform field conditions and slow-varying voltages. In the nonuniform field, previous results[1] obtained with a point-to-plane configuration indicated a difference in streamer formation between the positive and negative polarities. The study, then, was based on corona-current measurements, and the formation mechanism was not completely clear. To study this further, data were obtained simultaneously with high-speed camera, still camera, and photocell, as well as current measuring devices, thus giving a correlated and clear picture of the breakdown process of oil.

OIL HAS BEEN extensively used in electric apparatus as insulation for many years. During this time, a number of investigators[2–6] have postulated the breakdown mechanism of oil and the factors that affect its strength. However, those investigations were generally restricted to uniform field conditions and slow-varying voltages. Since impulse acceptance tests are becoming increasingly important in the electrical industry, and since the complex insulating structure in apparatus always contains some nonuniform field regions, an investigation of the impulse breakdown mechanism in a diverged field is therefore of practical interest. Komelkov[7] has used a point-to-plane arrangement to study the mechanism of impulse breakdown of liquids, including oil, and has obtained very interesting results. However, in his investigation a large value of resistance (of the order of 10^5 to 10^6 ohms) was placed in series with the test gap. The objection to such a technique is that it might seriously distort the mechanism being studied. These studies were made by the authors without insertion of resistances between the test gap and the output of the impulse circuit, which is a 1,100-micromicrofarad capacitor.

A point-to-plane test gap was chosen for several reasons. It permitted larger gap spacings without using excessively high voltages and was simple and needed little attention. Also, since nonuniform field regions are frequently found in electrical apparatus, a point-to-plane gap would more nearly duplicate these fields than uniform-field electrodes. In addition, even in some cases of quasiuniform-field regions in oil, the first breakdown streamer starting across the gap would distort the field of the gap, giving it nonuniform characteristics.

Definition of Terms

Since there has been a lack of accepted definitions for some of the terms that are used to describe certain phenomena discussed in this paper, an attempt is made to define such terminologies in the following, to facilitate a consistent usage.

CORONA is an intense ionization which modifies the dielectric characteristics of the oil or other insulation and causes light emission.

GLOW is a type of corona which in appearance is a visual glow of indefinite shape.

STREAMER OR CORONA-STREAMER is a type of corona having a distinct path of relatively long length.

INITIATING STREAMER is the first streamer in the oil, from which the subsequent formation of the corona mechanism begins.

STEP OR STEPPED STREAMERS denotes one or more streamers which follow after the initiating streamer in a manner of successive discrete steps.

GAP-BRIDGING STREAMER is a streamer which spans completely from one electrode to another electrode or to a barrier which arrests the further progress of the streamer in the oil gap.

RESTRIKE is any gap-bridging streamer which occurs after the first gap-bridging streamer.

DISCHARGE is a general term describing any of these phenomena.

Instrumentation

A sketch of the basic equipment used in these studies is shown on Fig. 1. A copper point was suspended 3 inches above a copper ground plane in good quality *10-C* oil, in a clear Lucite tank. A 3/8-inch insulating barrier was placed on the ground plane to limit the final breakdown current and light emitted, thereby permitting studies of the propagation of relatively faint corona streamers in oil by means of high-speed photography.

The impulse wave used in all these experiments was a 1.5x40-microsecond wave, because it is the standard wave for impulse testing. The crest voltages for the positive and negative polarities were 190 kv and 240 kv respectively. These voltages were just slightly higher than the critical breakdown voltage for a 3-inch gap in the oil used.

Three still cameras were arranged around and above the tank to obtain right angle views of each discharge. One of these still cameras was generally replaced by a high-speed *f2.3*-Boys camera,[8] which was used to obtain a continuous time sequence of the propagation of each discharge. This camera had a film velocity of about 1.6 mils per microsecond, allowing resolution of events about 1 microsecond apart in time when the film negatives were viewed in a microscope. The widths and light intensities of the discharge channels could then be determined with a microdensitometer. No shutter was used since the discharge itself supplied the necessary light. Since the film was moving at 1.6 mils per microsecond, the width of the discharge would appear on the film 1.6 mils wider than its actual width, if the discharge sustained for a microsecond of time; 3.2 mils wider for two microseconds, and so on. Therefore, the channel width or diameter reported here would be the apparent value, which is somewhat larger than the actual value.

As a supplemental method of studying the light emitted by the discharges, two photocells were placed at right angles around the Lucite tank and their outputs were connected through suitable amplifiers to a cathode-ray oscillograph. These photocells were heavily shielded against external radiation, and their viewing windows were screened by slit arrangements so that the upper photocell was

---

Paper 53-280, recommended by the AIEE Basic Sciences and Transformers Committees and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 17, 1953; made available for printing April 22, 1953.

T. W. LIAO and J. G. ANDERSON are with the General Electric Company, Pittsfield, Mass.

The authors wish to acknowledge the direction and encouragement given by J. H. Hagenguth and the assistance of G. H. Goedecke, J. Church, and A. I. Harju in carrying out the experiments.

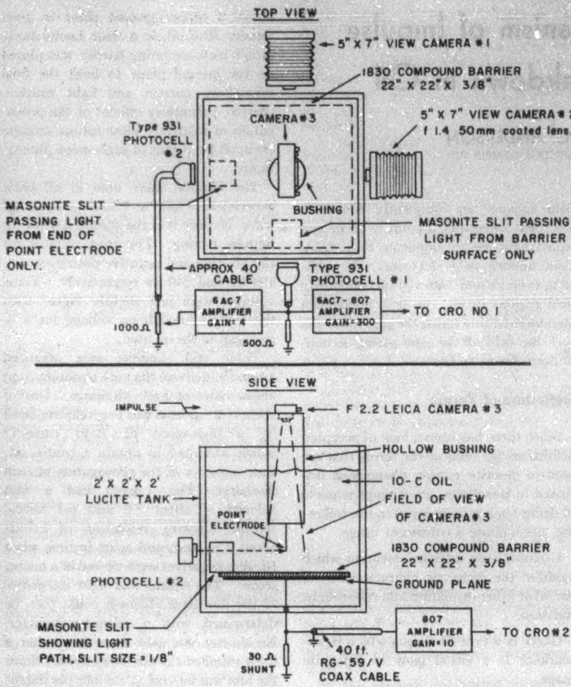

TOP VIEW
5" X 7" VIEW CAMERA # 1
1830 COMPOUND BARRIER
22" X 22" X 3/8"
5" X 7" VIEW CAMERA # 2
f 1.4 50mm coated lens.
CAMERA # 3
Type 931
PHOTOCELL
# 2
BUSHING
MASONITE SLIT
PASSING LIGHT
FROM END OF
POINT ELECTRODE
ONLY.
MASONITE SLIT PASSING
LIGHT FROM BARRIER
SURFACE ONLY
APPROX 40'
CABLE
TYPE 931
PHOTOCELL # 1
6AC7
AMPLIFIER
GAIN = 4
6AC7 - 807
AMPLIFIER
GAIN = 300
TO CRO. NO 1
1000Ω
500Ω
SIDE VIEW
IMPULSE
F 2.2 LEICA CAMERA # 3
HOLLOW BUSHING
10-C OIL
2' X 2' X 2'
LUCITE TANK
FIELD OF VIEW
OF CAMERA # 3
POINT
ELECTRODE
1830 COMPOUND BARRIER
22" X 22" X 3/8"
PHOTOCELL # 2
GROUND PLANE
807
AMPLIFIER
GAIN = 10
TO CRO # 2
MASONITE SLIT
SHOWING LIGHT
PATH. SLIT SIZE = 1/8"
30 Ω
SHUNT
40 ft.
RG-59/V
COAX CABLE

Fig. 1. Sketch showing instrumentation for study of impulse corona and breakdown in oil

sensitive only to light emitted near the point electrode, and the lower one only to light emitted at the insulating barrier or ground electrode surface. In this manner it was possible to record light pulses either near the point, or at the barrier, or both simultaneously, with two amplifiers and two oscillographs. In some cases, in order to save one amplifier and an oscillograph for measuring streamer currents, composite-light oscillograms were obtained by connecting the upper photocell to the first stage of a 3-stage amplifier and the lower photocell to the second stage of the same amplifier. Thus a streamer near the point would cause a deflection in one direction on the screen, and that near the barrier would result in a deflection of the reverse direction. However, the reverse deflections were many times smaller because they were only amplified by the two stages of the amplifier, unless the light intensity was very much stronger at the barrier.

In the electric circuit, current pulses were produced by the discharges in the oil. These current pulses were passed through a noninductive shunt, and the resulting voltages produced were amplified by a pulse amplifier and displayed on another cathode-ray oscillograph. The current pulses recorded by this method could then be compared with the light pulses picked up by the photocells to obtain a correlation between the current produced by a discharge and the light emitted from that discharge.

Streamer Propagation and Breakdown Mechanism

A set of typical photographs of the streamer propagation and oscillograms of the current and light pulses measured are presented in Figs. 2(A) and (B) for positive and negative polarities, respectively. Due to the loss of definition in the printing process, the high speed Boys photographs have been touched up slightly to show the fine details appearing on the negatives. The streamer propagation and breakdown mechanism to be described in the following are based on these results, and can be divided into four steps for each polarity.

Positive Point-to-Plane Gap, Fig. 2(A)

1. *Initiating Streamer:*

The breakdown starts at the point in the form of a glow, which lasts for a few microseconds. During this time, intense ionization occurs along a channel, developing into an initiating streamer.

2. *Successive Streamer Sropagation:*

The initiating streamer is followed by a series of stepped streamers with several microsecond time intervals between streamers, and with a channel diameter of approximately 10 mils. Each streamer is distinct from the preceding streamer, but all follow the same channel which progresses toward the ground plane. When the streamer branches into two parts, no appreciable change of its channel diameter or light intensity is observed. The streamers may further branch as they propagate toward the plane.

3. *Gap Bridged:*

The streamers eventually reach to and are arrested by the barrier on the ground plane, and the oil gap is then totally bridged by the streamers. If the barrier is removed, a breakdown will result.

4. *Restrikes:*

Successive streamers, each following the same path, may occur when the ground plane is covered by an insulating barrier, until the applied voltage drops to approximately half-crest value.

Negative Point-to-Plane Gap, Fig. 2(B)

1. *Initiating Streamer:*

The breakdown starts at the point in the form of a streamer with a channel diameter of approximately 35 mils and may span as much as half the gap within a microsecond of time.

2. *Successive Streamer Propagation:*

The initiating streamer is followed by a series of stepped streamers, with 1 to 5 microsecond time intervals and of much smaller channel diameter and light intensity. Unlike the positive stepped streamers, the tips of the negative stepped streamers appear to blend into a glow which progresses toward the ground plane. The number of the negative stepped streamers is, in general, two or three times that of the positive ones. There is no large branching like the positive streamer unless the initiating streamer has branches.

3. *Gap Bridged:*

When the glow reaches or approaches very close to the barrier on the ground plane, the gap-bridging streamer travels from the barrier back to the point electrode. The diameter of the streamer channel decreases toward the point, being approximately 50 mils near the barrier and about 25 mils at the point. The light intensity along the channel varied similarly. When a different material is used for the barrier, the light intensity and channel diameter of the gap-bridging streamer are also changed. It is shown in Fig. 3 that the gap bridging streamer with the glass barrier has a smaller channel. This dependence on the barrier material has not been found for the positive polarity. It is interesting to note that the transparency of the glass barrier has permitted so much light to be received by the lower photocell that the light pulse for the gap bridging streamer is actually reversed. If the barrier is removed, a breakdown will naturally result.

4. *Restrikes:*

When the ground plane is covered with an insulating barrier, one or more restrikes may occur with 10 to 15 microseconds interval between them, while the applied voltage drops to less than half-crest value. The light intensity is lower than that of the gap-bridging streamer but is much more uniform; the channel diameter was about 20 mils. The reversal of light pulses on the light oscillograms seems to indicate that there are ionizations on the barrier prior to and after the restrike, and in some cases, even prior to the gap-bridging streamer. The much stronger light intensity of the negative-polarity gap-bridging streamer near the barrier is also an indication of ionization on the barrier which provides a source of charges to flow in the channel of the gap-bridging streamer. No reversal of light pulses is found for the positive polarity, and as can be seen from Fig. 2(A), the light intensity along the channel of the positive gap-bridging streamer is much more uniform.

Comparison of Mechanisms of Lightning and Impulse Corona or Breakdown

Fig. 4 shows the remarkable similarity that exists between the mechanisms of natural lightning to open terrain, and the impulse critical breakdown of a negative point-to-plane gap in *10-C* oil, although the latter is on a much smaller scale.

Both phenomena appear to start with an electron avalanche rushing toward the ground plane from the negatively charged cloud or point. In the case of lightning, this avalanche may reach a distance of 150 feet and is known as a step leader.[9–11] For oil, the avalanche may reach a distance of 1 inch or so for a 3-inch gap having a crest voltage of 240 kv, and has been labeled an initiating streamer. This initial avalanche for both air and oil is then followed by a series of successive streamers which follow down the same channel, each streamer being longer than the previous one, until the ground plane is reached. The lightning streamers, or steps of the stepped-leader process, have

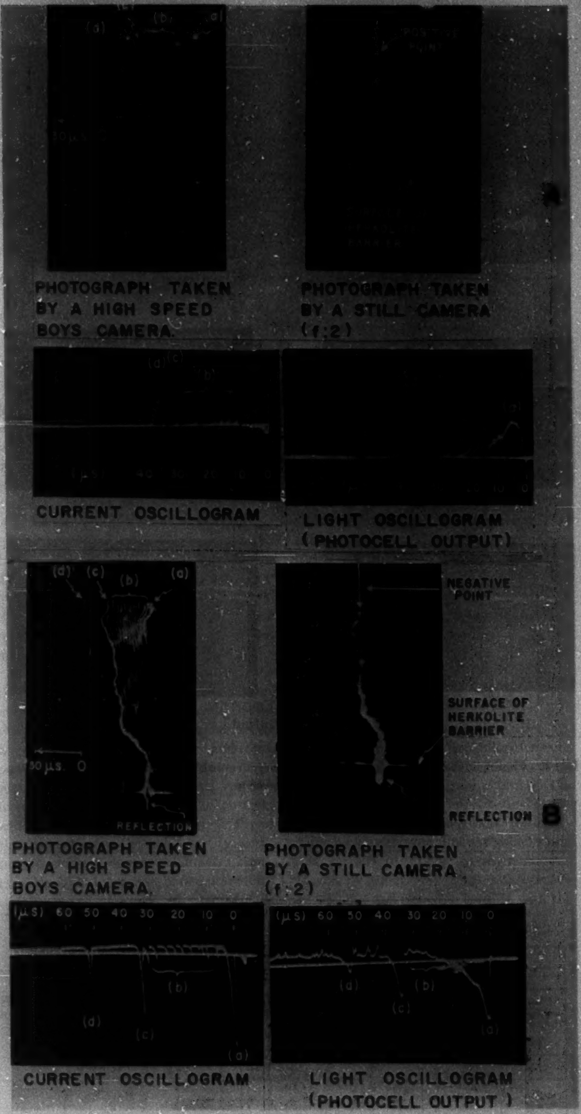

(d)
(b)
(a)
30 μs
PHOTOGRAPH TAKEN BY A HIGH SPEED BOYS CAMERA
PHOTOGRAPH TAKEN BY A STILL CAMERA (f:2)
(d)
(c)
(b)
(μs) 40 30 20 10 0
CURRENT OSCILLOGRAM
(a)
LIGHT OSCILLOGRAM (PHOTOCELL OUTPUT)
(d) (c) (b) (a)
30 μs 0
REFLECTION
NEGATIVE POINT
SURFACE OF HERKOLITE BARRIER
REFLECTION
B
PHOTOGRAPH TAKEN BY A HIGH SPEED BOYS CAMERA
PHOTOGRAPH TAKEN BY A STILL CAMERA (f:2)
(μs) 60 50 40 30 20 10 0
(b)
(d)
(c)
(a)
CURRENT OSCILLOGRAM
(μs) 60 50 40 30 20 10 0
(d)
(b)
(c)
(a)
LIGHT OSCILLOGRAM (PHOTOCELL OUTPUT)

Fig. 2. Photographs and oscillograms showing the mechanism of 1.5x40-microsecond impulse breakdown of a 3-inch point-to-plane gap in 10-C oil. (3/8-inch thick Herkolite barrier on plane)

A. Positive point-to-plane gap
B. Negative point-to-plane gap

(a) Initiating streamer
(b) Successive streamer propagation

(c) Gap bridged
(d) Restrike

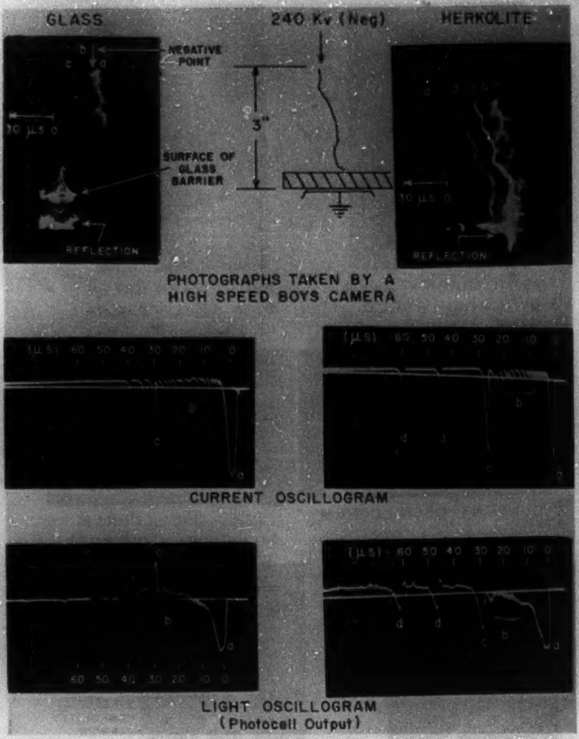

GLASS
240 Kv (Neg)
HERKOLITE
b
c
a
NEGATIVE POINT
30 µs 0
3"
SURFACE OF GLASS BARRIER
REFLECTION
d
30 µs 0
REFLECTION
PHOTOGRAPHS TAKEN BY A HIGH SPEED BOYS CAMERA
(µs) 60 50 40 30 20 10 0
CURRENT OSCILLOGRAM
LIGHT OSCILLOGRAM (Photocell Output)
60 50 40 30 20 10 0

Fig. 3. Photographs and oscillograms comparing negative-polarity oil corona streamers striking glass and Herkolite

an apparent velocity of about 150 feet per microsecond, but since the interval between steps may be 50 microseconds, the over-all propagation velocity of the stroke is much slower. The order of magnitude reported[9] is about 5 inches per microsecond. In oil, the velocity of the individual stepped streamers is unmeasurable with the equipment used, but should be over 4 inches per microsecond. The over-all velocity of propagation toward the ground plane, caused by delay of 1 to 5 microseconds between streamers, is of the order of 0.1 inch per microsecond. Thus, the breakdown process may take 30 microseconds to cross a 3-inch gap in oil.

It has been theorized[9–11] that the stepped leaders in lightning strokes are produced as an aftermath of the gradual progression to earth of a faint pilot streamer having a velocity equal to the over-all propagation velocity.

In Fig. 4 this pilot streamer is sketched in dashed lines. The high-speed photographs of breakdown in oil actually show an equivalent pilot streamer, progressing downward from the tip of the initiating streamer to the ground plane, in a manner very similar to that postulated for lightning. This equivalent pilot streamer has a velocity which is the same as the over-all propagation velocity of the breakdown, and appears as a faint hazy glow with occasional small branching.

As a stepped streamer for oil or a step leader for air reaches the ground plane, a heavy return stroke (or gap-bridging streamer) rushes upward from the ground plane to the cloud or point electrode as shown in Fig. 4. In the case of oil, when the ground plane is covered by insulation

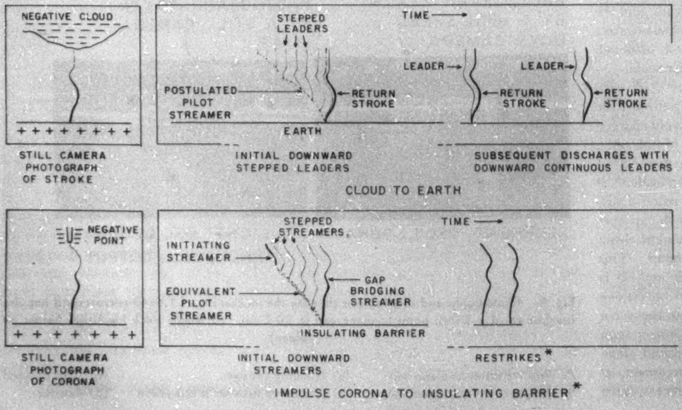

NEGATIVE CLOUD
+ + + + + + + + +
STILL CAMERA PHOTOGRAPH OF STROKE
STEPPED LEADERS
TIME →
POSTULATED PILOT STREAMER
RETURN STROKE
EARTH
INITIAL DOWNWARD STEPPED LEADERS
LEADER
RETURN STROKE
LEADER
RETURN STROKE
SUBSEQUENT DISCHARGES WITH DOWNWARD CONTINUOUS LEADERS
CLOUD TO EARTH
NEGATIVE POINT
+ + + + + + + +
STILL CAMERA PHOTOGRAPH OF CORONA
STEPPED STREAMERS
TIME →
INITIATING STREAMER
EQUIVALENT PILOT STREAMER
GAP BRIDGING STREAMER
INSULATING BARRIER
INITIAL DOWNWARD STREAMERS
RESTRIKES*
IMPULSE CORONA TO INSULATING BARRIER*

Fig. 4. Comparison of mechanisms of natural lightning from cloud to earth with negative-polarity point-to-plane impulse corona in oil

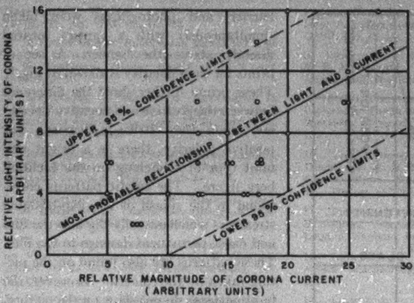


Fig. 5. Relationship between light intensity and current of positive-polarity streamers in oil

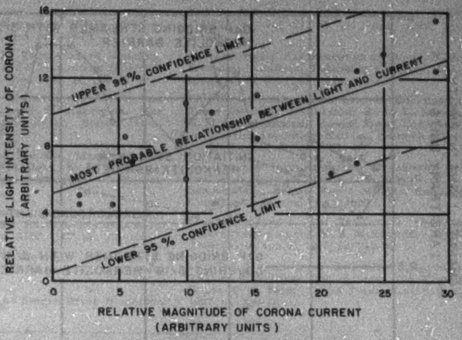


Fig. 6. Relationship between light intensity and current of negative-polarity streamers in oil

on its surface, the lower portion of the returning gap-bridging streamer is much more intense than the upper portion. This fact has occasionally been observed in lightning strokes to open terrain.[11] At intervals of 2 to 20 microseconds after the gap-bridging streamer occurs in oil, from the insulated ground plane, restrikes (subsequent gap-bridging streamers) are noted very similar to the multiple strokes occurring in lightning, except that their intervals are measured in microseconds, instead of tenths of seconds common to lightning observations.

However, if the ground plane in the point-to-plane oil gap is not protected by insulation, the gap will break down immediately when a stepped streamer reaches the ground plane, and no subsequent restrikes are possible.

Further Evaluation of Results

RELATION BETWEEN CORONA CURRENT AND LIGHT EMISSION

It has been found experimentally[12] that the light emitted from the channels of lightning strokes is, in general, proportional to the current in the stroke channel, but that the light has longer duration. Therefore, it is interesting to see whether this relation between light and current also holds for corona streamers in oil. The light pulses were recorded with a photocell and the current pulses were measured as described in "Instrumentation." About 30 current and light pulses were measured for each polarity of corona discharge, and their crest magnitudes were plotted in Figs. 5 and 6, with light as the ordinate and current as the abscissa, in arbitrary units. The points were somewhat scattered, but a straight-line relation appeared feasible when the data were analyzed statistically.[14] Such an analysis also gives confidence limits within which 95 per cent of all future points within this range can be expected to fall. Note, however, that the most probable curve in Fig. 6 does not pass near zero light intensity at zero current. This is due to the difference of the offset of the current and light oscillograms with respect of their zero lines. The oscillograms in Figs. 2(A) and (B) show that the time duration of the light from the streamers is considerably longer than the corresponding duration of current pulses, as is similarly found in gaseous discharges.

The variations of light intensity along the streamer channels were also determined. This was done by measuring the image density on the photographic negative with a microdensitometer[13] and computing the intensities from known density-intensity relationships. A large difference was found between the positive and negative polarities. Fig. 7 shows the fluctuation of the light density along the positive-streamer channel. However, the light intensities of the branching streamers and of the main streamer are about the same. But for the negative polarity, the difference of the light intensities between the initiating and the gap-bridging streamers is far more pronounced. The light intensity of the initiating streamer decreases gradually as the streamer leaves the point, whereas, for the gap-bridging streamers the intensities increased greatly near the barrier. The relative intensity and its rate of change is different for different barrier materials, see Fig. 8. The variation of the light intensity along the streamer channels for both polarities can also be seen from Figs. 2(A) and (B).

LICHTENBERG FIGURES

A number of Lichtenberg figures for both positive and negative polarities were obtained by laying a sheet of Kodak 5 by 7-inch Super-X photographic film on the barrier, where it was struck by corona streamers. The purpose was to determine how far the streamers splashed along the surface of a barrier when it was struck by them.[15,16] The emulsion side of the film was placed in contact with the

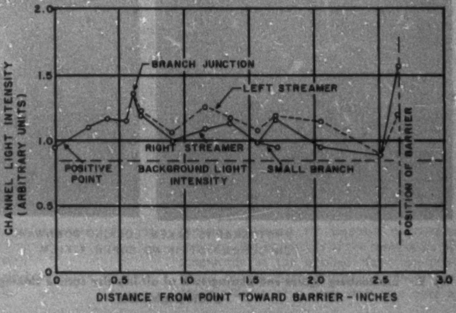


Fig. 7. Variation of light intensity along channel of positive - polarity streamer in oil

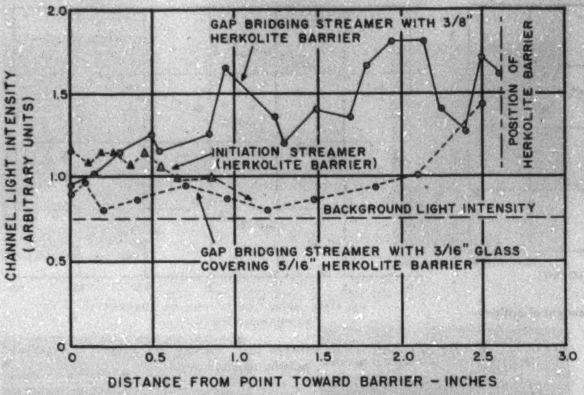
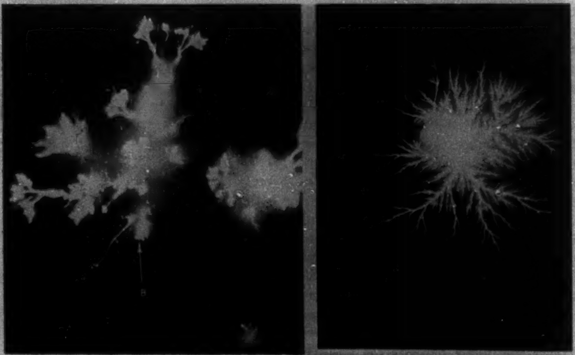
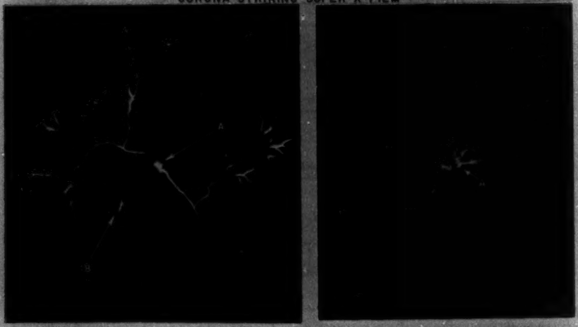


Fig. 8. Variation of light intensity along channels of negative-polarity streamer in oil using Herkolite and glass barriers



LICHTENBERG FIGURES PRODUCED BY CORONA STRIKING SUPER X FILM



PHOTOGRAPHS TAKEN LOOKING DOWNWARD ON CORONA STRIKING SUPER X FILM

Fig. 9. Lichtenberg figures and photographs of oil impulse corona striking back of a sheet of photographic film

barrier, and photographs were taken simultaneously with a camera located directly above the barrier. A set of prints for each polarity is shown in Fig. 9. These prints further show the difference between the positive and negative streamers, when they strike a barrier. For positive polarity, there is a bright and faint type of branching on the Lichtenberg figure. The bright patterns correspond to the splash of the gap-bridging streamers which actually dig into the film and cause permanent damage to the film. These patterns are also found on the picture taken by the camera. However, the faint patterns do not show on the picture, and they are probably caused by discharges between the barrier and the film. Nevertheless, the Lichtenberg figure gives a fairly good record of the area covered by the splash of the positive streamer. For the negative polarity, however, the Lichtenberg figure is caused only by the discharges between the barrier and the film and there appears to have been very little splash on the back of the film to expose the emulsion side. Only a very faint pattern was recorded by the camera and the film used for the Lichtenburg figure was not damaged by the negative discharge. The general treelike patterns are somewhat similar to those obtained in air,[17,18] especially for the positive polarity. For the negative polarity, the patterns produced by oil discharges are less diffused. This is probably due to smaller lateral diffusion of ions in each discharge channel; consequently the pattern is more distinct. It is expected that these figures can be used to measure approximately the barrier voltage, by comparing them with those produced with a point directly on the film, in oil, on top of the barrier.

Conclusions

1. The impulse breakdown of oil in the nonuniform field is a streamer mechanism.

2. A remarkable similarity to the stepped leader of natural lightning strokes exists for the negative streamer at the critical breakdown voltage.

3. The overall velocity of propagation of the streamer tip in oil is about 1/50 that of the stepped lightning leader.

4. Channel diameters in the oil gap ranged between 10 and 50 mils.

5. The maximum intensity of light emitted from a corona streamer channel is approximately proportional to the current in the electric circuit produced by the streamer.

References

1. Impulse Corona—Detection, Measurement of Intensity, and Damage Produced, J. H.

Hagenguth, T. W. Lao. *AIEE Transactions*, vol. 71, pt. III, Jan. 1952, pp. 461–65.

2. A Study of the Electrical Breakdown of Liquids by Means of Electro-optical Shutter, **H. W. Washburn.** *Physics*, New York, N. Y., vol. 4, 1933, p. 29.

3. Mechanism of Electrical Breakdown in Liquid Dielectrics, **L. Inge, A. Walther.** *Journal of Technical Physics*, Leningrad, USSR, vol. 1, no. 5–6, 1935, pp. 539–50.

4. Electric Breakdown in Insulating Liquids, **F. Kopplemann.** *Zeitschrift für Physik*, Berlin, Germany, vol. 16.5, 1935, pp. 125–41.

5. Dielectric Strength of Mineral Oils, **F. M. Clark.** *AIEE Transactions* (*Electrical Engineering*), vol. 54, 1935, pp. 50–55.

6. Breakdown Processes in Insulating Oils of Various Electric Strengths, **E. Conradi.** *Archiv für Elektrotechnik*, Berlin, Germany, vol. 30, 1936, pp. 677–82.

7. Mechanism of Dielectric Breakdown of Liquids, **V. S. Komelkov.** *Comptes Rendus* (*Doklody*), l'Academie des Sciences de l'URSS, Moscow, USSR, vol. 48, no. 4, 1945, pp. 268-70.

8. Lightning to the Empire State Building, **K. B. McEachron.** *AIEE Transactions* (*Electrical Engineering*), vol. 60, Sept. 1941, pp. 885–90.

9. Compendium of Meteorology (book), The Lightning Discharges, **J. H. Hagenguth.** The American Meteorological Society, Boston, Mass., 1951.

10. The Lightning Stroke: Mechanism of Discharge, **K. B. McEachron, W. A. McMorris.** *General Electric Review*, Schenectady, N. Y., vol. 39, 1936, pp. 487–96.

11. Progressive Lightning III, **D. J. Malan, H. Collins.** *Proceedings*, Royal Society of London, London, England, vol. A162, 1937, pp. 175–203.

12. Lightning, **J. W. Flowers.** ~~*General Electric*~~ *Review*, Schenectady, N. Y., vol. 47, ~~April 1944,~~ pp. 9–15.

13. Lightning to the Empire State Building—Part III, **J. H. Hagenguth, J. G. Anderson.** *AIEE Transactions*, vol. 71, pt. III, Aug. 1952, pp. 641–49.

14. Industrial Experimentation (book), **K. A. Brownlee.** Chemical Publishing Company, Inc., Brooklyn, N. Y., 1949, p. 65.

15. Edge Phenomena in Impulse Breakdown Tests on Solid Insulation, **E. L. C. Larson, G. M. L. Sommerman.** *Annual Report*, Conference on Electrical Insulation, National Research Council, Washington, D. C., 1950.

16. Impulse Ionization and Breakdown of Liquids in Uniform Fields, **C. J. Bute, G. M. L. Sommerman.** *Annual Report*, National Research Council, Washington, D. C., 1952.

17. Measurement of Transients by the Lichtenburg Figures, **K. B. McEachron.** *AIEE Transactions*, vol. 45, 1926, pp. 712–20.

18. The Measurement of Surge Voltages on Transmission Lines Due to Lightning, **Everett S. Lee, C. M. Foust.** *AIEE Transactions*, vol. 46, 1927, pp. 339–56.

◆

Discussion

G. M. L. Sommerman (Battelle Memorial Institute, Columbus, Ohio): The results obtained on the breakdown mechanism in oil are highly original and mark an important step forward in our understanding of the fundamental processes involved. This achievement has been possible because the authors have brought to bear simultaneously a number of recently developed measuring techniques on a basic arrangement of dielectrics and electrodes.

The writer, together with E. L. C. Larson and C. J. Bute (see refs. 15 and 16 of the paper), has measured streamer propagation velocities in liquids by the use of Lichtenberg figures obtained with impulse waves chopped at various time intervals. The propagation velocities were found to be much greater for positive streamers in chlorinated liquids than for either positive or negative streamers in hydrocarbon oils. This suggests that if the authors of the paper should try a chlorinated liquid in their setup and with their techniques, they would obtain results which differ in important respects from those reported for the hydrocarbon oil. Such results would give further information on the breakdown mechanism as it is related to the chemical structure of the medium.

Although, as the authors show, there is considerable similarity between their results with oil and the results of field investigations on natural lightning, the underlying mechanisms need not be similar in some cases. For example, in the case of restrikes, those observed in lightning may be caused by subsequent release of charge from portions of the cloud that are distant from that which released the initial charge. The restrikes found by the authors in oil must arise from a different cause. The following hypothesis is suggested: 1. After the initial discharge from a negative point, electrons are sprayed over the upper surface of the insulating barrier, thereby raising the potential of this surface to a high negative value. 2. The potential of the negative point continues to decrease because a $1^1/_2$x40 wave is applied to it. 3. After a lapse of several microseconds, the potential of the upper barrier surface is sufficiently more negative than that of the point that a reverse breakdown can occur along the original path where ions in the original discharge have not recombined completely. This hypothesis could be checked by repeating the experiment with a much longer wave, say $1^1/_2$x500, and observing if the restrikes in oil are eliminated or occur after much longer time intervals. Also, if the authors succeed in measuring the potential of the upper surface of the barrier from Lichtenberg-figure analyses, as suggested in the last paragraph of the paper, this might give information which would support the hypothesis.

B. F. J. Schonland (Bernard Price Institute for Geophysical Research, University of the Witwatersrand, Johannesburg, S. Africa): The authors are to be congratulated on a valuable piece of work. Although the pilot-streamer has been recorded as a preliminary to long-spark breakdown by Allibone[1] and by Hagenguth and Anderson (ref. 13 of the paper), it has never before been photographed either in lightning or the long spark as a streamer which creates the conditions for the establishment of a series of steps.

It is to be hoped that the work can be extended to breakdown in other liquid dielectrics, and perhaps in solids as well. This would afford most valuable information about the nature of the pilot process. In the writer's opinion, this process builds up a streamer mainly by photoionization in front of its advancing tip and not by collisional ionization. If this view is correct, further study in different media would add a great deal of knowledge to the general subject of impulsive breakdown.

Before seeing this paper, the writer had sent forward for publication a discussion of the pilot and step processes in lightning and the long spark, in the course of which a new explanation of the step process was put forward. Since it may be of value in estimating the possibilities of other media for step-breakdown studies, the main equation derived from this theory is quoted here

$$r/l + r^2/l^2 = 2X_2/X_1$$

where

r = the radius of the streamer (pilot or leader)

l = the length of a step

X_1 = the steady-field breakdown strength of the medium

X_2 = the average value of the field-strength required for long-spark breakdown over the whole gap.

For air $X_1/X_2 = 5$ hence $l/r = 6$, a value which agrees fairly well with observations on lightning and the long spark. In 10-*C* oil breakdown the quantity X_1, the writer is informed, is 500 kv per centimeter and X_2, from the authors' data, is 31.5 kv per centimeter. Hence l/r should be about 8. For negative streamers in oil $l = 0.25$ inch and $r = 0.025$ inch; hence $l/r = 10$, in reasonable agreement with the theory.

This theory of the step process requires that though the luminosity may vary abruptly, the flow of current should be fairly constant throughout the whole process, the pilot streamer carrying steadily forward the same charge per unit length as the good conducting channel behind it. The transient pulsations recorded on the authors' current oscillograms should thus be superimposed on a continuous background of current which flows in the intervals between pulsations, and which is of such magnitude that the quantity transported in the intervals between pulsations considerably exceeds that transported during the pulsation or step periods. The writer understands that the circuit arrangements used did not permit of measurement of this interstep current, but he would appreciate the authors' comments on this question and would ask whether in future work it will be possible to measure both the pulsating and the continuous components of the current. Such measurements would be of considerable interest but, as they will have to be made on the high-voltage side of the circuit, they may be difficult to achieve.

References

1. Development of the Spark Discharge, **T. E. Allibone.** *Nature*, London, England, vol. 161, 1948, pp. 970–71.

T. W. Liao and **J. G. Anderson**: We agree with Dr. Sommerman's hypothesis for the mechanism of restrikes in oil. Recent test results of the measurement of barrier potentials produced by impulse discharges indicate that potentials of the same order of magnitude as the applied wave are produced on the barrier surface. These potentials decay at a very slow rate over a period of many seconds. Hence, the point potential does decay at a much faster rate than the barrier potential, causing large reverse voltages to be produced across the gap. These reverse voltages may cause the restrike discharges to occur along the original preionized path.

The authors are pleased to learn that a reasonable agreement is obtained for the streamer propagation process in oil with Dr. Schonland's theory on the propagation processes in lightning and long sparks in air. At present no data are available on the interstep current. Measurements of such currents not only might provide a check of his theory applied to the oil medium but might also give some information on the mobility of ions in oil.

It is apparent from the written discussions and from some oral discussions with other investigators that a study of the streamer propagation processes in liquids is of general interest. The authors feel that this could be a very fruitful field of investigation for many workers, and hope that this paper may stimulate greater interest in the subject.

Pulse Response Characteristics of Rectangular-Hysteresis-Loop Ferromagnetic Materials

JOSEPH WYLEN
ASSOCIATE MEMBER AIEE

MAGNETIC CORES made with rectangular-hysteresis-loop ferromagnetic materials have recently gained prominence as binary elements in digital data-processing machines. The ability of the core to store binary information reliably and without power dissipation, and the ease with which the information can be read out or changed in a few microseconds make the core a useful and versatile component. Wherever they are applicable, magnetic-core circuits offer advantages over vacuum-tube circuits; the cores are rugged, have indefinite life, and require little power. Cores made with rectangular-hysteresis-loop ferromagnetic materials are being used in shift resisters,[1] in memory systems,[2,3] and in logical and control circuits.[4–6]

Because of the nonlinear nature of ferromagnetic materials the data on core characteristics that are needed for magnetic-circuit design must be obtained empirically. This paper defines some of the important properties of core materials and describes one method for measuring the quantities that are of interest in pulse applications. The test technique is described, and the results of tests made with several different core types are presented.

Definitions of Terms

Fig. 1 shows an idealized d-c hysteresis loop for a ferromagnetic material. The quantities that are of particular interest for this discussion are $\mathbf{B}_r$, the residual magnetization; $\mathbf{B}_m$, the maximum or tip magnetization; and $\mathbf{H}_c$, the coercivity or coercive force. Binary information can be stored in a core as the sense of the residual magnetization; $+\mathbf{B}_r$ is usually designated as binary "1"; $-\mathbf{B}_r$, binary "0."

Each system that employs magnetic cores depends on some unique set of characteristics, and hence places different specifications on the ferromagnetic material. Two properties of magnetic cores that are of general interest and importance are the squareness ratio and the switching time. The squareness ratio is the ratio of the residual magnetization $\mathbf{B}_r$ to the maximum magnetization $\mathbf{B}_m$. The switching time is here defined as the time required for the magnetization state to change from $-\mathbf{B}_r$ to $+\mathbf{B}_m$ when an unloaded core is subjected to a step of magnetizing force of amplitude $\mathbf{H}_m$. Other definitions of switching time, possibly more useful for circuit design, will not be treated here. Once a core has been switched from $-\mathbf{B}_r$ to $+\mathbf{B}_m$ its state has effectively been changed; the core will assume the $+\mathbf{B}_r$ state when the magnetizing force is removed.

Squareness Ratio

The importance of the squareness ratio $\mathbf{B}_r/\mathbf{B}_m$ is evident when one considers the effect of pulsing a core with a positive step of magnetizing force. Assume that a core having a cross-sectional area A and a hysteresis loop like that shown in Fig. 1 is initially in the $-\mathbf{B}_r$ state; when magnetizing force $\mathbf{H}_m$ is applied, the resulting change in flux will be $A(\Delta\mathbf{B}_s)$. If, however, the core had initially been in the $+\mathbf{B}_r$ state, the change will be $A(\Delta\mathbf{B}_n)$. Since the average voltage induced in a winding on the core is proportional to the total change in flux, the ratio $\Delta\mathbf{B}_s/\Delta\mathbf{B}_n$ is a measure of the signal-to-no-signal ratio that can be realized when the core is used in an information processing circuit. The relation between $\Delta\mathbf{B}_s/\Delta\mathbf{B}_n$ and the squareness ratio is stated in equation 1

$$\frac{\Delta\mathbf{B}_s}{\Delta\mathbf{B}_n}=\frac{1+\mathbf{B}_r/\mathbf{B}_m}{1-\mathbf{B}_r/\mathbf{B}_m} \tag{1}$$

The commercially available ferromagnetic materials that are of particular interest for use in digital computer circuits have maximum squareness ratios ranging from 0.90 to 0.96. In the saturation regions the slope of the hysteresis loop is nonzero, and the value of $\mathbf{B}_r/\mathbf{B}_m$ decreases, as the amplitude of the applied magnetizing force is increased. However, for the very rectangular loop materials, and for the values of $\mathbf{H}_m$ that are used, the change in squareness ratio is usually unimportant.

Switching Time

The cores that lend themselves for use in digital machines can readily be switched when unloaded, in times ranging from less than 1 microsecond to 10 or 20 microseconds. The switching time for a given material is a function of $\mathbf{H}_m$. For values of $\mathbf{H}_m$ much greater than $\mathbf{H}_c$ the switching time is approximately proportional to the quantity $1/(\mathbf{H}_m-\mathbf{H}_c)$. In applications where fast switching for small applied field is desired, a material having a small coercivity is needed.

The switching time of a metallic ferromagnetic core, fabricated from lamina-

---

Paper 53-286, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 16, 1953; made available for printing May 11, 1953.

JOSEPH WYLEN is with the Burroughs Corporation, Philadelphia, Pa.

The author wishes to acknowledge the assistance of his associates in the magnetic development sections of the Burroughs Research Laboratory, and in particular C. F. Devenny for his help and Isaac L. Auerbach for his interest and encouragement. The author is also indebted to the various manufacturers who supplied some of the magnetic cores and data that were used.

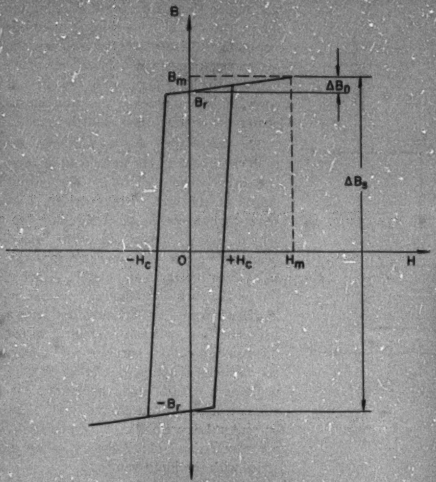


Fig. 1. Idealized d-c hysteresis loop

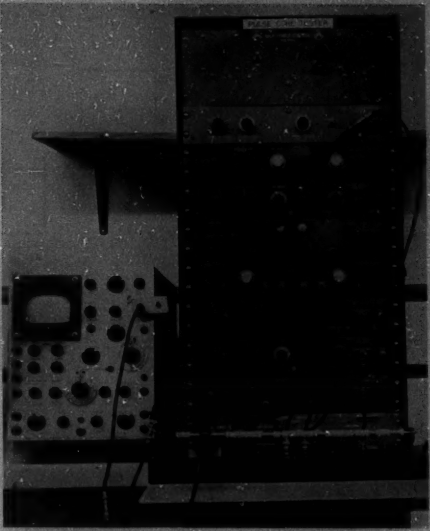


Fig. 3. Setup for toroidal core test. The equipment on the rack is used to generate the desired pulse pattern. The current drivers are not shown

tions thicker than 1/2 mil, is limited primarily by the eddy currents that flow within the core material during the flux change. The magnitude of the eddy currents is a function of the lamination thickness, the conductivity of the metal, and the magnitude of $\Delta\mathbf{B}_s$, see Fig. 1; eddy-current effects are minimized when all three of these factors are minimized. Hence, for a given value of $\mathbf{H}_m$, the fastest switching times are obtained with cores fabricated with ultrathin materials having small $\mathbf{H}_c$, small $\Delta\mathbf{B}_s$, and large resistivity. It is interesting to observe that although hysteresis loss in computer cores is negligible, fastest switching can be realized with the materials that have the smallest hysteresis loop.

The switching time of a core that is fabricated from laminations thinner than 1/2 mil cannot be predicted when only the eddy current effects are considered. Extensive investigation is being conducted to determine the factors that limit the switching speed of the very thin metallic tapes.

Ferromagnetic resonance and domain-wall relaxation and resonance, all of which contribute to core loss at high frequencies,[7] are possibly the phenomena that limit the switching times of the ferromagnetic ferrites. The eddy currents are not a significant factor; they are negligibly small because of the large resistivities of the ferrites. The coercivities of the commercially available rectangular-loop ferrites are more than ten times those of the useful metals; however, this fact alone is not sufficient to explain the delay in switching.

Materials

Recently developed ferromagnetic materials than can be switched rapidly and that have squareness ratios greater than 0.90 are of two types: grain-oriented ultrathin alloys, and ferrites. The two most important alloys, both available in strip as thin as 1/8 mil(0.000125 inch), are 48-per-cent nickel-iron and *4-79* Permalloy, a molybdenum-nickel-iron alloy. The processing, fabrication, and d-c hysteresis-loop characteristics of these two alloys have been fully described in a recent paper.[8] It will serve to mention that the alloys are cold reduced to the desired gauges with large percentage reductions at each step, slit into ribbon, usually 1/8-inch wide, coated with an insulating material, and wrapped onto ceramic toroidal bobbins. The free end of the outer wrap is tacked, either with a light spot weld or by other means. The assembly is annealed in a dry hydrogen atmosphere to maximize the squareness ratio of the magnetic material. The annealing process is critical, and must be carefully controlled if optimum results are to be achieved.

The coercivity of both of the rectangular loop alloys mentioned above increases as the gauge is made thinner, at least down to 1/8-mil gauge. This fact limits to some extent the advantages that can be realized by using thinner materials, as is illustrated in the experimental results presented below for 1/4-mil and 1/8-mil *4-79* Permalloy.

The maximum squareness ratio that can be obtained with a given alloy seems

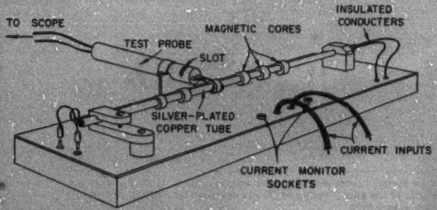


Fig. 2. Test equipment for toroidal magnetic cores

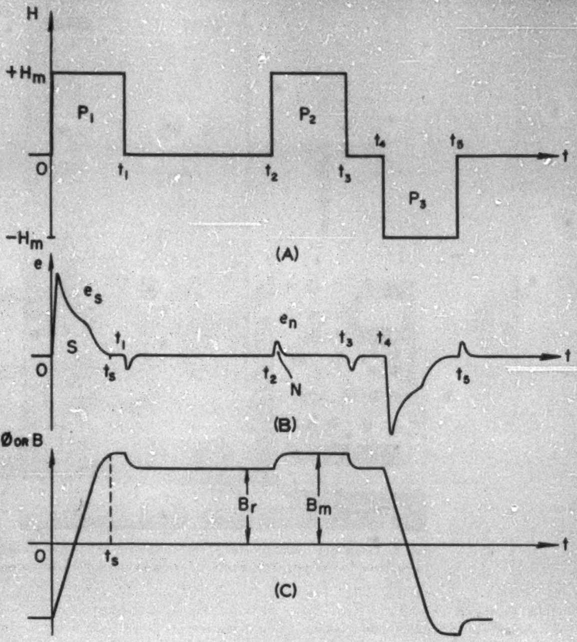


Fig. 4. Basic magnetic-core test

A. Applied magnetizing force for basic magnetic-core test
B. Typical core output voltage. The switching voltage is e_s; the nonswitching voltage is e_n
C. Magnetic flux or flux density

to be a function of the thickness to which it is rolled. In the case of the 48-per-cent nickel-iron, the squareness ratio decreases as the gauge is made thinner, while for the *4-79* Permalloy, the largest squareness ratio is achieved with the thinnest gauge.

The manufacturing processes and typical properties of the ferrites are described in the literature.[9,10] Finely powdered mixtures containing magnetite, various bivalent metals, and an organic binder are pressed into the desired shape and fired at high temperatures. The resulting compounds are ferromagnetic. The ferrite *MF 1118*, recently made available, has an exceptionally rectangular hysteresis loop, and is available in toroids having outside diameters as small as 0.06 inch.

Test Equipment

To facilitate the testing of large numbers of toroidal cores without recourse to winding each and every core, special test equipment was constructed. Because of its simplicity, the equipment is also useful for tests on single cores. Fig. 2 is a sketch showing the type of test jig that is used. The silver-plated copper tube serves both to support the cores, and as part of the single-turn output winding. Each of the insulated conductors that are threaded through the tube serves as a common single-turn input winding for all of the cores. One wire is used for applying positive magnetizing forces, the other, negative magnetizing forces. Because of the large permeability of the test cores, it is not necessary in practice to center the conductors within the tube.

The single-turn input windings are each connected as the plate load of their respective current drivers. The output stages of the two current drivers that were used to obtain all of the data presented below consist of eight *6CD6* beam-power tetrodes connected in parallel. Each driver can deliver current pulses that are continuously variable in amplitude from 0 to 7 amperes, and have a fixed rise time, without overshoot, of 0.4 microsecond. For purposes of safety, the cathodes of the output tubes are at a negative potential and the output circuits are nominally at ground level. The current pulses are observed by monitoring across precision 1-ohm resistors that are in series with the input windings.

The test probe consists of a two-tined fork, as shown in Fig. 2. Each tine makes electrical contact with the silvered tube on one side of the test core, and the tube surface and the two tines constitute a single-turn output loop. This loop encloses an area that is much greater than the cross-sectional area of the test core, and hence has induced in it by the air-coupled flux a relatively large voltage. The effect of this undesirable voltage is eliminated by including within the probe several turns that are oriented in such a way that the voltage induced in them exactly cancels the undesirable voltage. Since these reversed turns do not enclose the test core, they have no effect on the observed core output voltage. The reversed turns are inside the test probe, just behind the tines. An opening in the probe, see Fig. 2, permits adjustment of the size and orientation of the reversed windings; the self-positioning feature of

Table I. Description of Test Cores

All Metallic Ribbons are 1/8 Inch Wide

Test Core	Material	Nominal Ribbon Thickness, Mils	Mean Diameter, Inches	Number of Wraps	B_s,* Gausses	H_c,* Oersteds
1	48-per-cent nickel-iron	1.0	0.5	2	16,000	0.18
2	48-per-cent nickel-iron	1.0	0.5	2	16,000	0.18
3	48-per-cent nickel-iron	0.5	0.5	2	16,000	0.2
4	4–79 Permalloy	0.25	0.5	9	8,700	0.09
5	4–79 Permalloy	0.125	0.375	10	8,700	0.12
6†	MF 1118		0.58		2,350	1.57
7	4–79 Permalloy	0.5	0.375	16	8,700	0.07
8	4–79 Permalloy	0.25	0.375	32	8,700	0.09

* The data in these columns for metallic cores are taken from reference 6. B_s is the saturation flux density, presumably a true characteristic of the alloy; values of H_c are approximate and vary from core to core.
† Cross-sectional area = 0.0026 square inch.

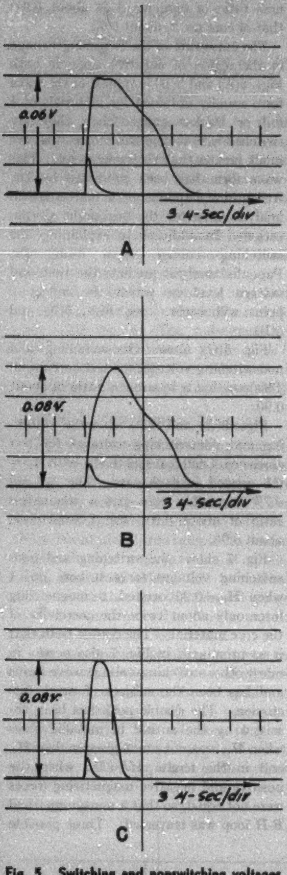


Fig. 5. Switching and nonswitching voltages for 48-per-cent nickel iron, $H_m=1.0$ oersted

A. Test core no. 1, 1 mil, $B_r/B_m=0.91$
B. Test core no. 2, 1 mil, $B_r/B_m=0.93$
C. Test core no. 3, 1/2 mil, $B_r/B_m=0.90$

the tines eliminates the necessity of frequent adjustments.

The laboratory setup used for core testing is shown in Fig. 3. The two large current drivers, normally at the bottom of the rack, are not shown. The pulse-control equipment above the test-jig shelf is used to generate the pulse train that triggers the drivers. The test probe is connected to the vertical input terminal of an oscilloscope either directly or, as noted in the following, through an integrating network.

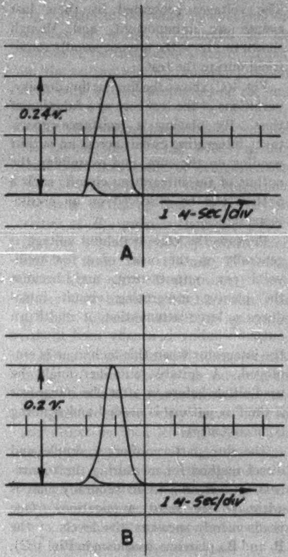


Fig. 6. Switching and nonswitching voltages for 4-79 Permalloy, $H_m=0.5$ oersted

A. Test core no. 4, 1/4 mil, $B_r/B_m=0.90$
B. Test core no. 5, 1/8 mil, $B_r/B_m=0.92$

Test Method

The pattern of applied magnetizing force that it used in the basic core test is shown in Fig. 4(A). If the test core is initially in the $-B_r$ state, see Fig. 1, pulse P_1 will switch the core to the $+B_m$ state. During the change, the switching voltage e_s, shown in Fig. 4(B), will be generated in the single-turn output winding. The duration of the switching voltage, t_s in Fig. 4(B), is the switching time of the core.

At the end of pulse P_1, the applied magnetizing force returns to zero and the core moves to the $+B_r$ state. The decrease in flux density is small if the core has a rectangular hysteresis loop, and a small negative voltage is induced in the output winding, as shown in Fig. 4(B), at time t_1.

A second positive pulse P_2 is applied to the core at time t_2, and the core is changed from the $+B_r$ state to the $+B_m$ state. The voltage e_n that is induced in the output winding by this change is the nonswitching voltage; it is always smaller in amplitude and shorter in duration than the switching, or signal, voltage. If the area under the e_s curve is designated S, and the area under the e_n curve is designated N, the area ratio is related to the squareness ratio of the core by equation 2

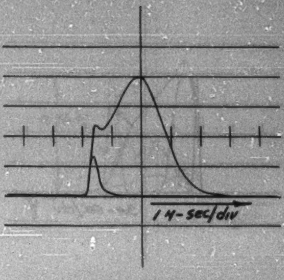


Fig. 7. Switching and nonswitching voltages of 1/4-mil 4-79 Permalloy core, test core no. 4, $H_m=0.2$ oersted. (Note the prominent double peak in the switching voltage)

$$\frac{S}{N}=\frac{1+B_r/B_m}{1-B_r/B_m} \tag{2}$$

Note that since the electromotive force induced in a single-turn winding is given by the rate of change of flux in the core, the area under a voltage curve is equal to the total change of flux that induced it. Hence, equations 1 and 2 are equivalent.

For test purposes, the oscilloscope sweep is triggered just before pulses P_1 and P_2 are applied, and the voltages e_s and e_n are observed superimposed on the oscilloscope screen. The switching time t_s can be measured directly if the sweep speed is properly calibrated. And since equation 3 can be obtained from equation 2 by slight manipulation

$$\frac{B_r}{B_m}=\frac{S-N}{S+N} \tag{3}$$

the superimposed voltage pattern offers a means for measuring the squareness

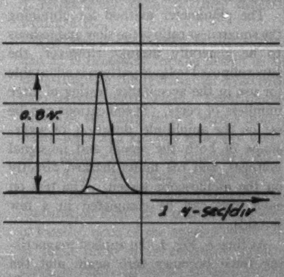


Fig. 8. Switching and nonswitching voltages of ferrite MF-1118, test core no. 6, $H_m=3.0$ oersteds, $B_r/B_m=0.95$

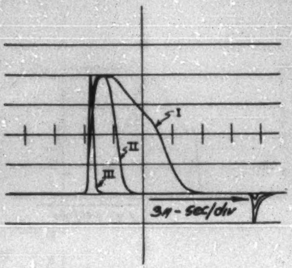


Fig. 9. Switching voltages compared for $H_m = 1.0$ oersted

- I. 1-mil 48-per-cent nickel-iron, test core no. 1
- II. 1/2-mil 48-per-cent nickel-iron, test core no. 3
- III. 1/4-mil 4–79 Permalloy, test core no. 4

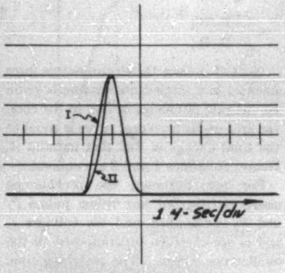


Fig. 10. The switching voltages of 1/4- and 1/8-mil 4–79 Permalloy compared for $H_m =$ 0.5 oersted

- I. Test core no. 4
- II. Test core no. 5

ratio—a planimeter can be used to measure both (S−N) and (S+N) directly from a copy of the oscilloscope pattern.

The planimeter method for obtaining the squareness ratio is too slow and tedious to be generally useful. However, the superimposed voltage pattern is adaptable for use in the acceptance testing of large numbers of cores. Specifications can be set on the amplitude, duration, or waveshape of both the signal and no-signal voltages, and the limits marked directly on the oscilloscope screen. The test on each core can be completed in a few seconds.

At time t_4, Fig. 4, the applied magnetizing force becomes zero again and the core returns to the $+\mathbf{B}_r$ state. Pulse P_3 causes the core to be reset to the $-\mathbf{B}_r$ state, after which the cycle is repeated. The voltages generated by these last events are unimportant, and, though shown in Fig. 4(B), are not usually monitored during the test.

Fig. 4(C) shows the flux, or flux density, variations of the test core as a function of time. By placing a resistance-capacitance integrating circuit across an output winding on the core, and examining the output of the integrating circuit, such a pattern can be observed on an oscilloscope screen.

Because the core switching voltage is generally on the order of a few millivolts per output turn, and because the passive integrating circuit introduces a large attenuation, a multiturn output winding is usually used to drive the integrator when this technique is employed. A suitable amplifier might be used either before or after the integrator if the flux pattern is needed and winding is inconvenient.

The flux pattern offers a simple and direct method for measuring the squareness ratio $\mathbf{B}_r/\mathbf{B}_m$ to an accuracy that is adequate for most applications. One needs merely measure the levels of the $\mathbf{B}_r$ and $\mathbf{B}_m$ plateaus, as shown in Fig. 4(C), and take their ratio.

Pulse Response Characteristics

Listed in Table I are the descriptions of the magnetic cores that were used for obtaining the test data presented in this section. The metallic cores were fabricated on ceramic bobbins in the manner previously described. All of the cores were selected from commercial lots as typical samples, but are not intended to represent any mathematically defined average.

The oscillograms in Figs. 5 through 18 are tracings of photographs.

Switching and Nonswitching Voltages

Fig. 5(A), shows, superimposed, the switching and nonswitching voltages for test core no. 1. Fig. 5(B) shows the same two voltages for test core no. 2. Both cores were taken from the same lot; the dissimilarity in the switching voltage waveforms is typical of the variations that occur. Variations in switching time can most likely be attributed to nonuniform ribbon thickness, to differences in coercivity, or to poor insulation between wraps. The variations in waveform might be caused by differences in the fabrication, such as unlike tensions during wrapping or nonuniform insulation thickness between wraps. The squareness ratio of core no. 1 is about 0.91; that of core no. 2, about 0.93.

The curvature of the switching voltage in the region of negative slope in both Figs. 5(A) and 5(B) is typical of the wave form usually obtained for materials of 1 mil or thicker gauge when they are switched with magnetizing forces that are much greater than their coercivity. This wave form has been predicted by Dr. Athanasios Papoulis in a mathematical analysis done for the Burroughs Corporation. In addition to explaining the switching voltage wave form, Dr. Papoulis' analysis predicts the time and voltage level at which the inflection point will occur. See Figs. 5(A) and 5(B).

Fig. 5(C) shows the switching and nonswitching voltages for test core no. 3. This core has a squareness ratio of about 0.90.

Figs. 6(A) and 6(B) show the switching and nonswitching voltages for test cores nos. 4 and 5, respectively, with $\mathbf{H}_m =$ 0.5 oersted in each case. The 1/4-mil *4-79* Permalloy core has a squareness ratio of about 0.90; the 1/8-mil core, about 0.92.

Fig. 7 shows the switching and nonswitching voltages for test core no. 4 when $\mathbf{H}_m = 0.20$ oersted, a magnetizing force only about twice the coercivity of the core material. The double peak that is so prominent in Fig. 7 also occurs in many other switching-voltage wave forms and has been the subject of much discussion. The double peak has been observed by the author in metallic cores when $\mathbf{H}_m$ was not much greater than $\mathbf{H}_c$, and in the ferrite *MF-1118* when the positive and negative magnetizing forces were not equal, so that a nonsymmetrical **B-H** loop was traversed. Three possible

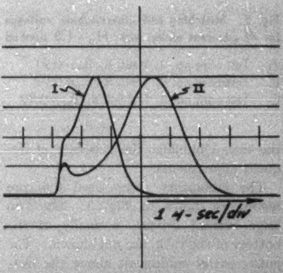


Fig. 11. The switching voltages of 1/4- and 1/8-mil 4–79 Permalloy compared for $H_m =$ 0.25 oersted

- I. Test core no. 4
- II. Test core no. 5

causes for the double peak are:

1. Nucleation of the domains;
2. The sudden change in slope from the flat to the rising portion of the B-H loop;
3. Magnetostrictive effects.

Another possible reason could be a difference in coercivity between the regions near the surface of the core and the regions within the core, if the difference is large and if the surface coercivity is the greater of the two. Further study of the double-peak phenomenon, now being conducted at various laboratories, should lead to a satisfactory explanation.

Fig. 8 shows the switching and nonswitching voltages for the ferrite sample, test core no. 6. The pulse-measured squareness ratio of cores made of *MF-1118* is about 0.95 when the core is switched on the saturation hysteresis loop.

SWITCHING TIME VERSUS THICKNESS

Fig. 9 compares the switching times for test cores nos. 1, 3, and 4, all being switched by 1.0-oersted magnetizing pulses. The switching voltage for 1/4-mil *4-79* Permalloy is compared, in Fig. 9, with the two 48-per-cent nickel-iron samples to demonstrate that for a given value of applied field, fastest switching will be realized with a core that has the thinnest gauge, the lowest maximum flux density, and the smallest coercivity, see Table I.

The oscilloscope amplifier gain was adjusted to make the amplitudes of all three waveforms equal for the composite in Fig. 9 (as well as in the two composites which follow). This was done to give a better comparison of wave form and switching time. The amplitude of the switching voltage can be varied by changing the number of wraps of material on the bobbin or the number of turns of copper on the output winding, and is not dependent only on the properties of the material.

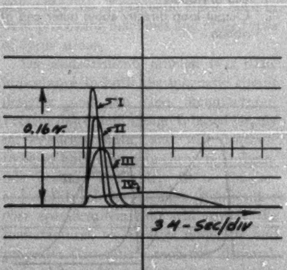


Fig. 12. The switching voltages of 1/2-mil 48-per-cent nickel-iron, test core no. 3, for four different values of applied magnetizing force

I. $H_m=1.5$ oersteds
II. $H_m=1.25$ oersteds
III. $H_m=1.0$ oersted
IV. $H_m=0.5$ oersted

Fig. 10 compares the switching voltages for 1/4-mil and 1/8-mil *4-79* Permalloy (test cores nos. 4 and 5) being switched by 0.5-oersted pulses. Both cores have the same switching time; it is only for values of $\mathbf{H}_m$ greater than 0.5 oersted that the 1/8-mil Permalloy switches more rapidly than the 1/4-mil Permalloy. Note that if eddy currents were the only factor that limited switching time, the 1/8-mil material would switch most rapidly.

Although the two wave forms pictured in Fig. 10 reach their peak values in the same time (about one microsecond), and decrease together, they have different curvatures on their leading edges. This may be dependent on the 0.4-microsecond rise time of the magnetizing pulse, although it is interesting to note that the switching voltages continue to increase for some time after the applied magnetizing fields reach peak value.

The rise time of the magnetizing force that is applied to a fast switching core does influence the pulse-response characteristics of the core. However, no data showing the effects of varying the rise time have been included in this paper.

Fig. 11 compares again the switching voltages for test cores nos. 4 and 5, this time for $\mathbf{H}_m=0.25$ oersted. The 1/8-mil material, although of thinner gauge than the 1/4-mil material, switches much more slowly than the latter. In part, this is probably because of the larger coercivity of the 1/8-mil Permalloy, which results in a larger value for the quantity $1/(\mathbf{H}_m-\mathbf{H}_c)$. Since the double peak is very prominent in the switching voltage of the 1/8-mil sample, it is probable that a satisfactory theory concerning the double peak would explain further the slower switching of the thinner gauge material.

SWITCHING TIME VERSUS APPLIED MAGNETIZING FORCE

Fig. 12 is a composite that shows the four switching voltages that result for four different values of magnetizing force applied to test core no. 3. Wave form IV demonstrates that the double-peaked switching voltage is also obtained with the 48-per-cent nickel-iron cores when $\mathbf{H}_m$ is small. The areas under each of the four voltage wave forms are approximately equal, since each results from approximately the same total change of flux. Figs. 13 and 14, also composites, show the switching voltages of test cores nos. 4 and 6, respectively, each for three different values of $\mathbf{H}_m$.

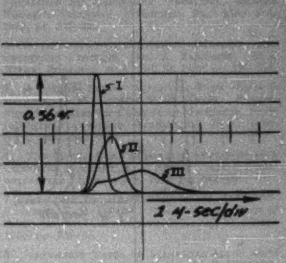


Fig. 13. The switching voltages of 1/4-mil 4-79 Permalloy, test core no. 4, for three different values of applied magnetizing force

I. $H_m=1.0$ oersted
II. $H_m=0.5$ oersted
III. $H_m=0.25$ oersted

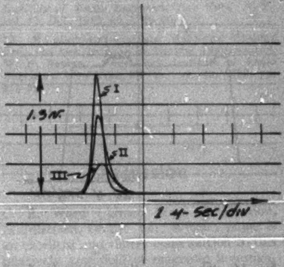


Fig. 14. The switching voltages of the ferrite MF-1118, test core no. 6, for three different values of applied magnetizing force. (The decreased area under wave form III indicates traversal of a minor loop)

I. $H_m=4.0$ oersteds
II. $H_m=3.0$ oersteds
III. $H_m=2.0$ oersteds

The type of data shown in Figs. 12, 13, and 14 is necessary for the design of circuits that use the rectangular-hysteresis-loop ferromagnetic cores. Point-plotted curves of switching time t_s versus magnetizing force $\mathbf{H}_m$ are not generally useful, as is made evident by consideration of Fig. 14, the switching voltages for the ferrite sample. Though traversing the "saturation loop" of a material is spoken of loosely, none of the applied fields used in these tests, or in computer applications, is really sufficient to saturate the cores. Each different value of applied $\mathbf{H}_m$ causes the core to traverse a slightly different loop, and each loop has different

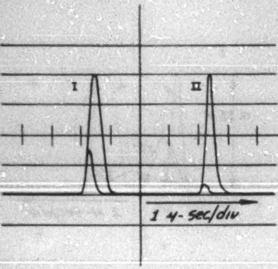


Fig. 15. The switching and nonswitching voltages of 1/2-mil and 1/4-mil 4–79 Permalloy for $H_m=2.0$ oersteds

I. Test core no. 7, $B_r/B_m=0.7$
II. Test core no. 8, $B_r/B_m=0.9$

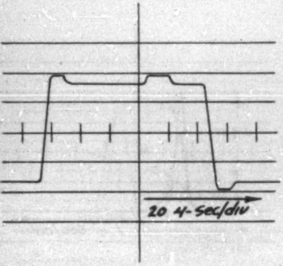


Fig. 16. Flux versus time pattern for 1-mil 48-per-cent nickel-iron, test core no. 1. The squareness ratio is about 0.91

values of $\mathbf{H}_c$, $\mathbf{B}_r$, and $\mathbf{B}_m$. The switching times for $\mathbf{H}_m=2.0$ oersteds and $\mathbf{H}_m=3.0$ oersteds in Fig. 14 are equal, but it is obvious that the area under wave form III, representing the total change in flux for $\mathbf{H}_m=2.0$ oersteds, is much less than the area under wave form II, which represents the total change in flux for $\mathbf{H}_m=3.0$ oersteds. A curve of t_s versus $\mathbf{H}_m$ would indicate the same switching time for the two different values of $\mathbf{H}_m$, but would not, of itself, indicate that two different hysteresis loops were involved. Hence, the importance of actual wave forms.

SQUARENESS AND GAUGE

The composite in Fig. 15 is included to illustrate that the squareness ratio of 4-79 Permalloy increases as the gauge is decreased. The change in squareness ratio between 1/2-mil and 1/4-mil material is quite marked, and comparison of the switching and nonswitching voltages of the two samples makes it immediately apparent. The 1/2-mil sample has a squareness ratio of about 0.7; the 1/4-mil sample, about 0.90.

As stated earlier, the squareness ratio of 48-per-cent nickel-iron decreases as the gauge is decreased. Figs. 5(B) and 5(C) might be compared to illustrate this, although it is not obvious from simple inspection.

FLUX VERSUS TIME

Fig. 16 shows the flux versus time pattern for test core no. 1. Such a pattern depicts, within the limits of the accuracy of the integrating network, the total change in flux that has taken place within the core between any times t_a and t_b. However, it does not picture the manner in which the magnetic field actually penetrates the core. The flux pattern in Fig. 16 indicates that the core has a squareness ratio of about 0.91.

EFFECT OF LOAD

Throughout this discussion, only unloaded cores have been treated; the single-turn output winding has negligible effect on the core characteristics. However, the operational characteristics of a core in a magnetic circuit are dependent largely on the load it must drive. In a typical application, the magnetic shift register,[1] the load consists of a similar core, several diodes, leakage inductance, and distributed capacity.

The treatment of the effects of various loads on core characteristics is beyond the scope of this paper. However, it is important to mention that the distributed capacity of a multiturn winding can have significant effect on the switching time, even when the winding is not connected to any external load. A 100-turn winding was placed on test core no. 1 in order to obtained the flux pattern shown in Fig. 16. Although it is not evident in the figure, a 20-per-cent increase in switching time resulted. A 300-turn distributed winding on a core like test core no. 4 can actually triple the observed switching time.

GENERAL

An interesting effect that can be observed with the ribbon-wrapped cores is illustrated in Fig. 17. Wave forms I are obtained when the single-turn output loop encloses the region diametrically opposite the spot-welded area of the ribbon; wave forms II are obtained when the output loop is directly about the spot-welded area. The difference in squareness ratio is immediately apparent. This effect is probably the result of flux leakage in the vicinity of the end of the ribbon, but an exact cause has not been assigned. Improved core operation can be realized in circuits where squareness ratio is important by placing the output winding in regions removed from the ribbon end.

Squareness ratios that approach unity are achieved in the alloys by subjecting the ferromagnetic materials to strain-relieving anneals in a dry hydrogen atmosphere at soak temperatures as high as 2,000 degrees Fahrenheit. Any event which will upset the results of the anneal will decrease the squareness ratio. Fig. 18 illustrates dramatically the effect of pressing a fingernail firmly on the metal of test core no. 1. Wave forms I are the "before"; wave forms II, the "after."

Fig. 18 is not intended to indicate that the metallic cores are too sensitive to be useful. The stress used to obtain these results was much greater than that which a core would receive in almost any realistic condition. The metal is normally protected by the ceramic bobbin,

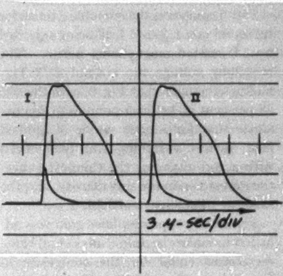


Fig. 17. Composite oscillogram showing effect of varying the position of a single turn output loop (test core no. 1)

I. Output loop diametrically opposite outer end of ribbon
II. Output loop directly about outer end of ribbon

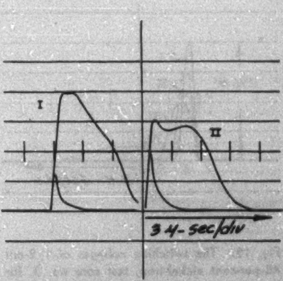


Fig. 18. Composite oscillogram showing effect of strain on switching and nonswitching voltages of 1-mil 48-per-cent nickel-iron, test core no. 1

I. Before stressing
II. After stressing

Table II. Values of Switching Time and Squareness Ratio of Commercially Available Rectangular Loop Materials

Material	Nominal Thickness, Mils	H_m, Oersteds	t_s, Microseconds	B_r/B_m*
48-per-cent nickel-iron	1	1.5	7 to 9	0.93
		1.25	8 to 10	
		1.0	11 to 12	
48-per-cent nickel-iron	0.5	1.5	2	
		1.25	3	
		1.0	4 to 6	0.90
		0.5	13 to 15	
4–79 Permalloy	0.25	1.0	0.8 to 1.0	
		0.5	1.8 to 2.2	0.90
		0.25	3 to 4	
		0.2	4 to 6	
4–79 Permalloy	0.125	0.5	1.8 to 2.2	0.92
		0.25	5 to 6	
MF 1118		4.0	1.0 to 1.5	
		3.0	1.5 to 2.0	0.95
		2.0†	1.5 to 2.0	

* Changes in squareness ratio are negligible for the small changes in H_m that are considered.

† A minor hysteresis loop is traversed.

and proper packaging techniques can be used to provide whatever additional protection is required.

It should be mentioned that the phenomena illustrated in Figs. 17 and 18 are most prominent with cores that have only one or two wraps of metal, and that the use of five or more wraps on a bobbin serves to temper the effects of both rotation and externally applied stress.

Conclusions

Rectangular-hysteresis-loop ferromagnetic materials have already proved their value in shift registers and in magnetic memories. Research programs currently being conducted in several laboratories throughout the country give promise of still wider use of magnetic cores, both in data-processing machines and other types of applications.

The measurements problem is of basic importance to both the magnetic circuit designer and the core manufacturer. Present day circuit design is based largely on empirical data, and will continue to be so until adequate theoretical methods become available. It is the responsibility of the engineer to formulate core specifications that will permit the manufacturer to develop materials and processes that satisfy each new demand. Only an assemblage of meaningful reliable data will enable the designer to select and use magnetic cores wisely and the manufacturer to control his product. It has been the purpose of this paper to indicate the kind of data that is useful. The magnetic-core test equipment that has been described is equally adaptable for use as a laboratory instrument or as a semiautomatized production tester, and hence can serve both the designer and the manufacturer.

Typical pulse-response characteristics of some of the commercially available rectangular hysteresis loop materials have been presented and discussed. The significant data, which were presented in the form of oscillograms within the paper, are summarized in Table II.

A simple and direct technique for measuring the pulse response characteristics of magnetic cores has been described. The test method is an extension of work performed at the Burroughs Research Laboratory and reported in an earlier paper.[11] The switching time and squareness ratio were selected for presentation because of their fundamental significance. However, the same principle can be used for more specialized tests if the proper pattern of applied magnetizing force is selected. An important example is the test that is used to select cores for the coincident-current memory.[12]

Pulse testing might also be employed advantageously for cores that are intended for use in magnetic amplifiers; those skilled in the magnetic-amplifier art can perhaps attach new interpretations to the data and techniques that have been reported here.

References

1. STATIC MAGNETIC STORAGE AND DELAY LINE, **An Wang, Way Dong Woo.** *Journal of Applied Physics*, New York, N. Y., vol. 21, 1950, pp. 49–54.

2. DIGITAL INFORMATION STORAGE IN THREE DIMENSIONS USING MAGNETIC CORES, **J. W. Forrester.** *Journal of Applied Physics*, New York, N. Y., vol. 22, 1951, pp. 44–48.

3. A STATIC MAGNETIC MEMORY SYSTEM FOR THE ENIAC, **Isaac L. Auerbach.** *Proceedings*, Association for Computing Machinery, Richard Rimbach Associates, Pittsburgh, Pa., May 2–3, 1952, pp. 213–22.

4. STATIC MAGNETIC MATRIX MEMORY AND SWITCHING CIRCUITS, **Jan. A. Rajchman.** *RCA Review*, Princeton, N. J., vol. 13, 1952, pp. 183–201.

5. A HIGH SPEED MAGNETIC-CORE OUTPUT PRINTER, **B. M. Gordon, R. N. Nicola.** *Proceedings*, Association for Computing Machinery, Sauls Lithograph Company, Washington, D. C., Sept. 8–10, 1952, pp. 6–12.

6. THE SINGLE-CORE MAGNETIC AMPLIFIER AS A COMPUTER ELEMENT, **R. A. Ramey.** *AIEE Transactions*, vol. 71, pt. I, Jan. 1953, pp. 442–46.

7. ANALYSIS OF MEASUREMENTS ON MAGNETIC FERRITES, **C. D. Owens.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 41, March 1953, pp. 359–65.

8. ULTRATHIN TAPES OF MAGNETIC ALLOY WITH RECTANGULAR HYSTERESIS LOOPS, **M. F. Littmann.** *AIEE Transactions*, vol. 71, pt. 1, July 1952, pp. 220–23.

9. MAGNETIC FERRITES, **C. L. Snyder, E. Alberts-Schoenberg, H. A. Goldsmith.** *Electrical Manufacturing*, New York, N. Y., Dec. 1949, pp. 86–91.

10. NEW DEVELOPMENTS IN FERROMAGNETIC MATERIALS (book), **J. L. Snoek.** Elsevier Publishing Company, New York, N. Y., 1947.

11. BEHAVIOR OF RECTANGULAR HYSTERESIS LOOP MAGNETIC MATERIALS UNDER CURRENT PULSE CONDITIONS, **E. A. Sands.** *Proceedings*, Institute of Radio Engineers, New York, N. Y. vol. 40, Oct. 1952, pp. 1246–50.

12. A COINCIDENT-CURRENT MAGNETIC MEMORY CELL FOR THE STORAGE OF DIGITAL INFORMATION, **W. N. Papian.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 40, 1952, pp. 475–78.

◆

Discussion

C. E. Ward (Armco Steel Corporation, Middletown, Ohio): Mr. Wylen's paper is an excellent and timely contribution to this relatively new field of magnetic investigations.

As a supplier of thin ferromagnetic tapes, my company is endeavoring to evaluate the effect of core dimensions on the magnetic properties of ultrathin materials as a means of arriving at true material properties. We have found that one of the major factors that must be taken into consideration is the degree of coupling that exists between the primary and secondary windings, with respect to leakage flux produced by the initial rapid change in exciting current.

The double peak in voltage that is referred to by Mr. Wylen as occurring at values of H_m slightly greater than the coercive force appears to be the result of the combined effect of air flux and iron flux. Normally, no initial voltage spike is observed for values of H_m considerably greater than H_c because the iron flux is changing rapidly and the voltage produced is large enough to mask the effect of air flux on the observed voltage pattern. However, when H_m is small the magnitude of the air flux voltage may appear to be quite large compared to the voltage produced by the iron flux. This is particularly true if the cross-sectional area of the iron is small relative to the area of the search coil and if the position of the windings are such that there is considerable leakage-flux coupling between the primary and secondary.

Our experiments have shown that the leakage field will be distorted by the presence of magnetic material. Therefore, it is possible that the compensation for air flux provided in the test probe of Mr. Wylen's equipment is correct with no sample in the field, and incorrect under actual test conditions.

The factors which influence the coupling between the primary and secondary winding with respect to leakage flux are 1. The placement of windings with respect to each other; 2. The ratio of cross-sectional area of iron to cross-sectional area of air encompassed by the search coil; 3. The dimension of the core bobbin.

Figs. 19(A) and (B) illustrate the type of voltage patterns produced by varying the coupling between primary and secondary windings from a maximum to a minimum for $H_m = 0.5$ oersted. Figs. 20(A) and (B) were obtained under the same test conditions, except that $H_m = 1$ oersted.

It is, of course, difficult to eliminate the effects of circuitry in magnetic measurements with pulse excitation. However, it is obvious that care must be exercised in both measuring techniques and commercial applications if the true characteristics of the core materials are to be utilized effectively.

It is expected that the results of our study of the effects of bobbin size, number of wraps of material, and placement of windings will be published in the near future.

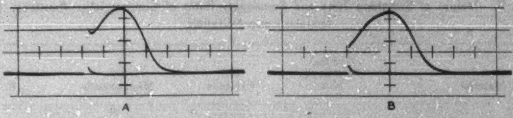


Fig. 19. Effect on switching and nonswitching voltages of air-flux coupling between primary and secondary windings, 1/2-mil Orthonik, $H_m = 0.5$ oersted

A. Maximum coupling
$V = 0.1$ volt per division
$T = 5$ microseconds per division
$B_r/B_m = 0.84$

B. Minimum coupling
$V = 0.1$ volt per division
$T = 5$ microseconds per division
$B_r/B_m = 0.97$

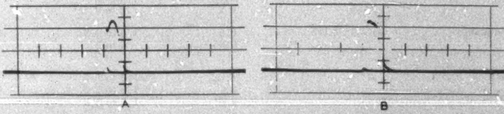


Fig. 20. Effect on switching and nonswitching voltages of air-flux coupling between primary and secondary windings, 1/2-mil Orthonik, $H_m = 1$ oersted

A. Maximum coupling
$V = 0.5$ volt per division
$T = 5$ microseconds per division
$B_r/B_m = 0.79$

B. Minimum coupling
$V = 0.5$ volt per division
$T = 5$ microseconds per division
$B_r/B_m = 0.96$

The Permeability of Silicon-Iron at Very Low Flux Densities

EBERHARD BOTH
NONMEMBER AIEE

THE initial permeability of magnetic materials is a characteristic of great theoretical as well as practical importance. By definition, the initial permeability is equal to the ratio of the peak value of the flux density to the peak value of the exciting magnetizing field, when the latter approaches the value of zero. Due to the difficulties encountered in measuring the true initial permeability, the term is frequently applied in a rather loose manner to values which have been determined at peak flux density values of 5 gausses or more, or which have been obtained by extrapolating permeability-versus-field strength curves to the zero value of field strength.

This practice is apt to cause considerable confusion, when the characteristics of various materials are to be compared and evaluated for applications for which the permeability at extremely small field strength is of critical importance as, for instance, in transformers, chokes, and filters used in wire communication.

In this paper the term "initial permeability" is used for values which were determined at field strength of the order of 10 microoersteds corresponding to peak flux density values of the order of 5 to 20 milligausses.

In his classical study of the behavior of iron and steel under the operation of feeble magnetic forces, which was published in 1887, Lord Rayleigh found[1] that for very small fields the permeability-versus-field strength curve can be represented by the equation $\mu = \mu_i + \nu H$ and that the same constant which expresses the increase in permeability with increasing field strength appears as a factor in the equation for the hysteresis loss $W_h = \nu H^3/3\pi$ (ergs per centimeter per cycle).

The significance of this relationship, which is known as Rayleigh's law, for the characteristics required of magnetic materials operating in very small fields is obvious: Since hysteresis losses not only drain energy from the sometimes very feeble signal but also distort it by introducing harmonics, the ideal communication transformer material would have a ν value of zero, that is, its permeability would be absolutely constant over the operating field range and its hysteresis loss would be zero.

Due to the work of H. Jordan[2] and V. E. Legg[3] the hysteresis constant has been universally accepted as one of the quality criteria of magnetic powder cores used in Pupin coils, filters, and other components of long-distance telephone circuits.

In 1943, the German communication industry adopted a standard specification[4] in which a modification of the Rayleigh constant was used to determine the acceptability of nickel-iron and silicon-iron alloys for use in communication transformers.

The requirements for the permeability and the hysteresis loss constants as stipulated in the German specification are

---

Paper 53-259, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 3, 1953; made available for printing May 1, 1953.

EBERHARD BOTH is consultant with the Signal Corps Engineering Laboratories, Fort Monmouth, N. J.

The author acknowledges his indebtedness to his associates, R. J. Franklin, D. M. Grossman, M. C. Hankel and W. Schlosser, who prepared, treated, and tested the specimens, and who are to be credited with the intelligent analysis and resolution of many unexpected difficulties encountered in the course of this work.

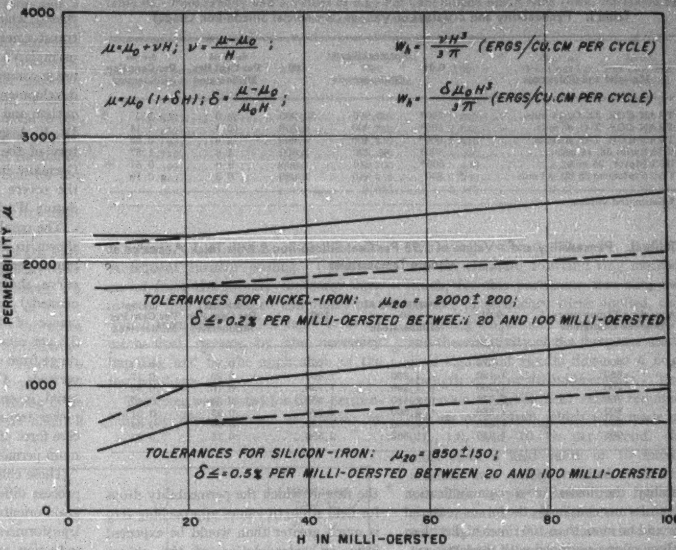


Fig. 1. Tolerances for permeability and hysteresis loss constants required for communication transformer materials according to the German Specification DIN E 41301

shown in Fig. 1. In the upper part of the figure, the Rayleigh equations for μ and W_h have been written in the original form using the constant ν, and in the modified form used in the specification, where the constant denoting the increase in permeability is called δ and is equal to ν/μ_0.

The two pairs of solid lines shown in the lower part of the figure represent the upper and lower permeability limits for nickel-iron alloys containing up to 40 per cent nickel and for silicon-iron alloys. For the former, the permeability tolerance at 20 millioersteds is 2000±200, for the latter 850±150; the maximum permissible δ-values are 0.2 and 0.5 per cent per millioersted respectively. For practical reasons the δ-values according to this specification are determined for the field strength range between 20 and 100 millioersteds.

It is beyond the scope of this paper to compare the performance characteristics of the nickel-iron materials commonly used in communication transformers in this country with their German counterparts, although it may be stated in passing that the adoption of the lower nickel contents prevalent in German materials would yield substantial savings in nickel without loss in performance characteristics.

For the problem at hand, however, a comparison of the representative silicon-iron grades presently available in this country with the German material Trafoperm *25N1* proves very informative. Such a comparison is made in Fig. 2, where data taken from current catalogues of the Armco Steel Corporation and the United States Steel Corporation have been plotted to the same scale as that used in Fig. 1.

It is at once apparent that in all the materials shown the change in permeability with field strength is many times higher than that exhibited by the Trafoperm *25N1* material. The δ-values calculated from these curves show a ratio of more than 100 to 1 between the topmost material TRAN-COR 3X-O and Trafoperm *25N1*, which means, of course, that under the assumption of equal perme-

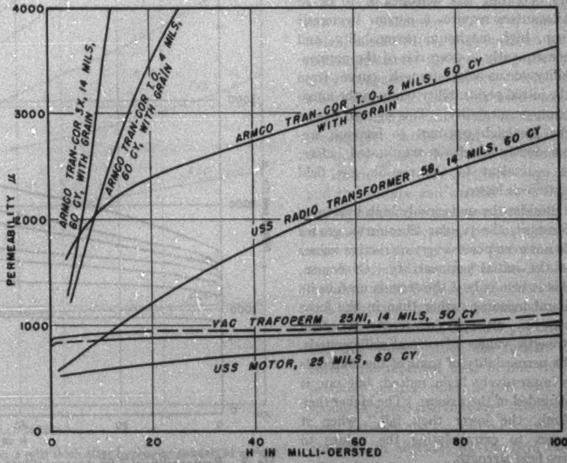


Fig. 2. Permeability of various commercial silicon-iron materials

Table I. Permeability and δ-Values of Various Commercial Silicon-Iron Grades

Material and Thickness	Permeability at 0.01 Millioersteds	Permeability at 20 Millioersteds	Permeability at 100 Millioersteds	$\delta_{0.01-100}$ Per Cent Per Millioersted	δ_{20-100} Per Cent Per Millioersted
TRAN-COR 3X-O, 14 mils	750*	5,900	30,000	39.0	5.1
TRAN-COR T-O, 4 mils	500*	3,500	6,700	12.4	1.14
TRAN-COR T-O, 2 mils	1,000*	2,400	3,600	2.6	0.63
USS Radio 58, 14 mils	500*	1,200	2,800	4.6	1.67
USS Motor, 25 mils	500*	640	900	0.8	0.51
VAC Trafoferm 25 N1, 14 mils	830	940	1,080	0.3	0.19

* Estimated values.

Table II. Permeability and δ-Values of 3.25 Per Cent Silicon-Iron 2 Mils Thick Annealed at Various Temperatures

Annealing Temperature, Degrees Centigrade	Permeability at 0.01 Millioersteds	Permeability at 20 Millioersteds	Permeability at 100 Millioersteds	$\delta_{0.01-100}$ Per Cent Per Millioersted	δ_{20-100} Per Cent Per Millioersted
600	870	915	990	0.14	0.10
650	1,120	1,230	1,400	0.25	0.17
700	1,200	1,390	1,750	0.46	0.32
750	1,270	1,480	1,950	0.54	0.40
800	1,380	1,690	2,520	0.83	0.61
850	1,520	1,900	2,950	0.94	0.69
900	1,330	1,600	2,320	0.75	0.56

ability the losses in a communication transformer containing the former material would be more than 100 times higher than those of one equipped with the latter material.

This statement will probably come as a shock to the manufacturers of grain-oriented steel, who are justly proud of the fact that their material has the lowest power losses of any of its kind produced anywhere. Paradoxically, however, this is the very reason for the fact that the losses at very low fields are particularly high, because achieving low power losses for operating flux densities of 10 or 15 kilogausses requires a narrow hysteresis loop, high maximum permeability, and consequently a steep rise of the permeability-versus-field strength curve from the initial permeability value to the maximum permeability. This in turn causes the Rayleigh constant to increase tremendously which, as was stated before, is equivalent to the rise in low field hysteresis losses.

Besides the undesirably high hysteresis constant, the regular silicon-iron grades do not even possess very attractive values of the initial permeability. Of course, this is true only if the term is used in its literal meaning rather than in the loose way so often encountered. At field strength values of several millioersteds, the permeability of some of the materials is impressively high, indeed, but one is reminded of the axiom: "The higher they climb, the lower they fall," when it comes to extrapolating the values to zero field strength.

It is noteworthy that in all the curves the rate at which the permeability drops for field strength values approaching zero is much greater than would be expected from an extrapolation of the almost linear portion of the curves at slightly higher fields. This behavior is characteristic of silicon-iron, and studies by J. L. Snoek,[5] F. Pawlek,[6] and others make it probable that the anomaly is due to very small impurities of carbon, nitrogen, oxygen, or possibly other elements contained in the metal.

In contrast to the materials just discussed, which might be classified as power transformer materials, the German Trafoperm *25N1* was exclusively developed for use in communication transformers. This development was undertaken by the author and his associates, most notably Dr. F. Assmus, at the Research Laboratory of the Vacuumschmelze in Hanau, Germany in 1941. It was prompted by the severe nickel shortage in Germany during World War II.

The notable features of the material, as shown in Fig. 2, are its relatively high initial permeability of about 830 in the curve shown (measured at 10 microoersteds) and the relatively small increase of the permeability, averaging 0.3 per cent per millioersted for the field range from 10 microoersteds to 100 millioersteds. On the other hand, the material is entirely unsuitable for use in power transformers, since it has a coercive force of 0.8 to 1 oersted and a maximum permeability of only about 6,000.

These characteristics are achieved by a process differing from the production of grain-oriented silicon-iron used in power transformers by a higher degree of cold reduction and a final anneal at a lower temperature.

In Table I the characteristics of the materials represented in Fig. 2 are summarized numerically.

The work reported here was done recently at the Signal Corps Engineering Laboratories. The material used in the study was obtained from the Allegheny Ludlum Steel Corporation in the form

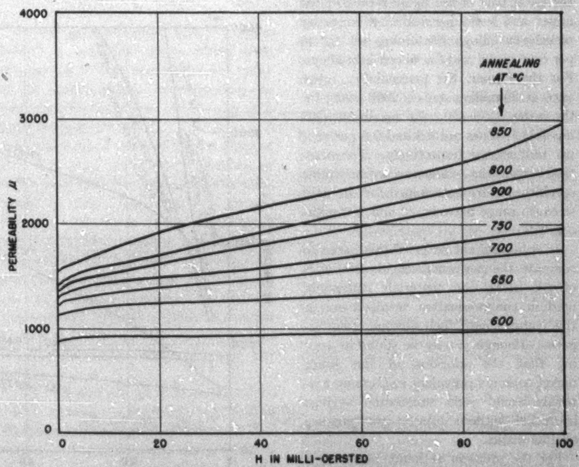


Fig. 3. Permeability of 3.25 per cent silicon-iron 2 mils thick after hydrogen anneals at various temperatures

of hard-rolled tape 0.5 inch wide and 5 mils, as well as 2 mils, in thickness. It had been produced from a fully annealed and oriented strip of 14 mils thickness, the normal power transformer material of the firm known as Silectron, by cold-rolling without intermediate anneal to the final thickness, thus resulting in a cold reduction of about 64 per cent for the 5-mil tape and of about 86 per cent for the 2-mil tape.

The material was coated with dry magnesium oxide powder for the purpose of insulating the adjacent layers of the tape and then wound to form toroidal cores with an inside diameter of about 1.5 inches and an outside diameter of about 2.0 inches. Individual series of cores were annealed in electrolytic hydrogen at peak temperatures ranging from 600 to 900 degrees centigrade in steps of 50 degrees centigrade. The annealing cycle consisted of a heating period of about 2 or 3 hours, a holding time of 2 hours, and subsequent furnace cooling. Upon removal from the furnace the cores were inserted into ring-shaped containers machined from resin-bonded paper to avoid mechanical stresses by the necessary handling and by the application of the test windings.

Table III. Permeability and δ-Values of 3.25 Per Cent Silicon-Iron 5 Mils Thick, Annealed at Various Temperatures

Annealing Temperature, Degrees Centigrade	Permeability at 0.01 Millioersteds	Permeability at 20 Millioersteds	Permeability at 100 Millioersteds	$\delta_{0.01-100}$ Per Cent Per Millioersted	δ_{20-100} Per Cent Per Millioersted
600	770	860	1,020	0.33	0.23
650	1,460	1,760	2,340	0.60	0.41
700	1,720	2,230	3,310	0.93	0.61
750	1,680	2,380	4,460	1.65	1.10
800	1,850	2,700	5,450	1.95	1.27
850	1,920	2,950	6,280	2.27	1.41
900	1,950	3,050	6,760	2.47	1.52

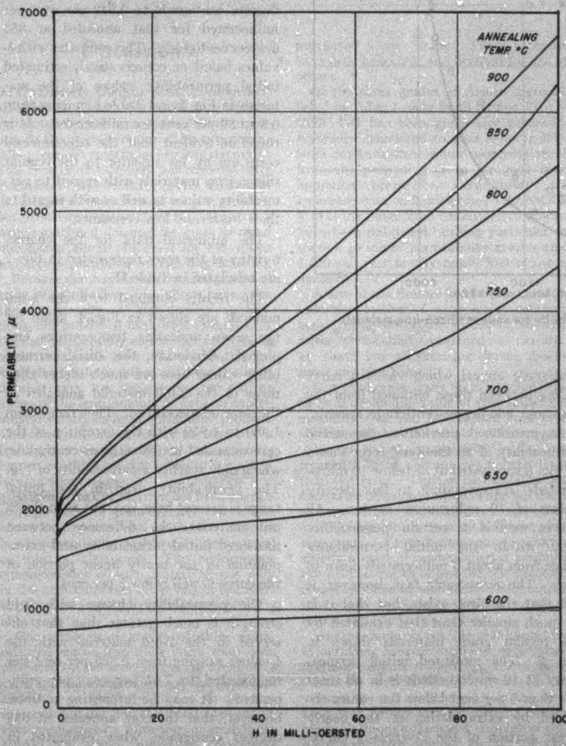


Fig. 4. Permeability of 3.25 per cent silicon-iron 5 mils thick after hydrogen anneals at various temperatures

The cores were tested for their permeability by direct current as well as by alternating current. The d-c method was the standard Rowland ring method which, for this particular core size, the number of secondary turns applied, the prevailing magnitude of the permeability, and the sensitivity of the fluxmeter used, gave significant results down to a field strength of 5 or 10 millioersteds. The extension to lower field values required the use of a-c tests which were made at 0.1, 1.0, and 10 kc per second. In most cases a field value of 10 microoersteds was safely reached with 10 kc. This is particularly true for the 2-mil material, where significant readings could be taken in some cases close to, or even slightly below, 1 microoersted. For the 5-mil material the limiting frequency proved to be 1.0 kc per second for several of the cores annealed in the upper temperature range, probably due to inadequate insulation or relatively high permeability values. Even in the worst case, however, the readings taken at 50 microoersteds are considered reliable.

The circuit for the a-c tests consisted of a Hewlett Packard oscillator model *650A* as the source, the test windings, and two Ballantine *VTVM*'s, one of which, model *310A*, was operated as a milliamperemeter for measuring the primary magnetizing current, while the other one, model 300, was used—sometimes in conjunction with an external preamplifier, model *220A*—to read the secondary voltage developed by the core. All instruments were carefully calibrated and the voltage and current wave shape was monitored for excessive distortion by means of an oscilloscope, but, as was to be expected at the very low flux densities, the wave shape remained as closely sinusoidal as could be judged by this check. It is estimated that the absolute accuracy of the results is well within 5 per cent of the true values, while the relative accuracy for repeated tests of one core is probably better than 2 per cent. It should be noted also that very close agreement between the a-c and d-c values was observed in all cases for the field range where both test methods overlapped.

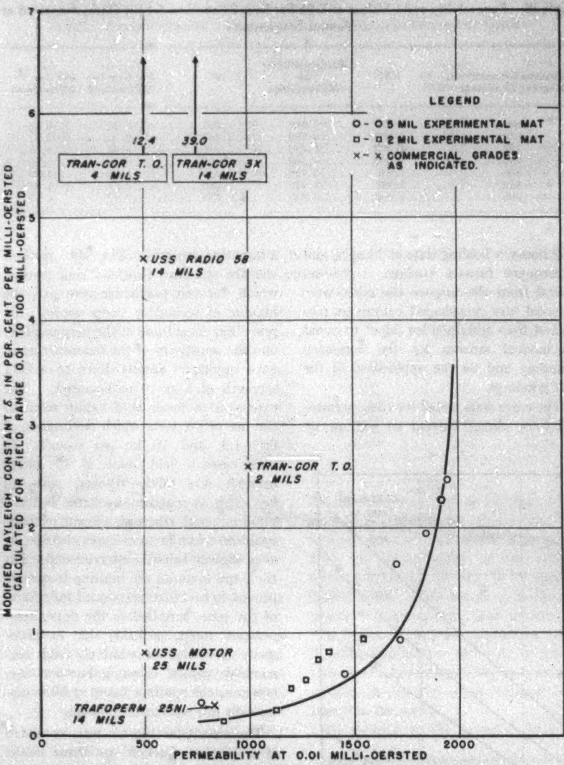


Fig. 5. Rayleigh constant versus initial permeability for various silicon-iron materials

Before the test, the cores were demagnetized by continuously reversing a direct current of slowly decreasing amplitude. This process was performed in two steps for which the number of primary turns was changed from 35 to 1, which allowed a very nearly continuous field reduction from a peak of 25 oersteds to below 10 microoersteds. This method of demagnetization proved to give much more consistent results than various a-c procedures which had been tried previously.

The results of the annealing experiments obtained with the 2-mil tape are plotted in Fig. 3, where the permeability-versus-field strength curves resulting from the seven different peak annealing temperatures are shown. Generally speaking, the initial permeability and the slope of the curves are seen to increase with increasing annealing temperature, with the exception of the 900-degree-centigrade anneal which yields a curve falling between those obtained from the 750- and 800-degree-centigrade anneals.

As mentioned previously, the initial permeability of all the 2-mil cores shown in Fig. 3 was tested at or below 10 microoersteds, corresponding to flux density values of 10 milligausses or less. All curves exhibit a certain permeability drop within the initial permeability range from about 3 millioersteds down to zero. The remarkable fact, however, is not that this drop exists, but that it is so much smaller than that exhibited by the regular grade materials shown in Fig. 2. The measured initial permeability at 10 microoersteds is in all cases less than 5 per cent below the values obtained by extrapolation of the nearly linear portion of the μ-versus-H curve. For the regular grades, accurate figures for this discrepancy cannot be determined, since information on the true initial permeability values is lacking, but it appears safe to say that it amounts to 50 per cent or possibly more in several cases.

The initial permeability values of the experimental cores range from 870 to 1,520 at 10 microoersteds. They may not appear to be spectacularly high, but it is believed that they are the highest ever reported for cores of 2-mil silicon-iron prepared from commercial material. They do not reach the values for nickel-iron alloys, but they approach them closely enough to warrant a study of their potential usefulness as a substitute material.

The variation of the permeability with field strength is much smaller than that of any of the regular materials. The δ-values calculated for the field range from 10 microoersteds to 100 millioersteds range from 0.138 per cent per millioersted for the material annealed at 600 degrees centigrade to 0.942 per cent per millioersted for that annealed at 850 degrees centigrade. This compares with δ-values based on conservatively estimated initial permeability values of the materials in Fig. 2 and Table I ranging from 0.8 to 30 per cent per millioersted. It is therefore evident that the experimental cores are by far superior to the regular silicon-iron materials with regard to permeability values as well as with regard to their hysteresis loss constants.

The numerical data for the characteristics of the cores represented in Fig. 2 are tabulated in Table II.

The results obtained with the 5-mil material are shown in Fig. 4, again for the seven annealing temperatures employed. Obviously, the initial permeability values here are much higher than those of the 2-mil material annealed at the same temperatures. They range from 1,460 to 1,950 with the exception of the core annealed at 600 degrees centigrade, which only reached a permeability of 770. The permeability drop in the initial range is present here, too, but as in the 2-mil material, the difference between measured initial permeability and extrapolation of the nearly linear portion of the curve is well below 5 per cent.

The permeability change with field strength is much greater than that observed in the 2-mil material with the δ-values ranging from 0.325 per cent per millioersted to 2.47 per cent per millioersted. It may be interesting to note, however, that the core annealed at 650 degrees centigrade, when evaluated in terms of the German standard specification, comes close to the permeability range of the nickel-iron class, having a permeability of 1,760 at 20 millioersteds, where a minimum of 1,800 is specified for nickel-iron, and that the δ-value for the field range from 20 to 100 millioersteds is well within the limit specified for the silicon-iron class, being 0.411 per cent per millioersted instead of 0.5 per cent per millioersted as specified.

It is possible that the poorer δ-values of the 5-mil material are at least partially due to the smaller amount of cold-reduction which this material has received, as was mentioned earlier.

Table III contains the summary of the numerical results of the cores represented in Fig. 4.

Fig. 5 represents an over-all review of the data discussed before. The δ-values of the regular grade silicon-iron as well as those of the experimental 2-mil and 5-mil material are plotted versus the initial permeability. It demonstrates once more that the regular grade materials possess extremely high hysteresis loss constants and poor initial permeability, while the experimental material is considerably better on both counts. From the tentative curve drawn to connect the optimum results of the experimental cores, it can be concluded that initial permeability values of 1,500 can be achieved, if the δ-value is limited to 0.5 per cent per millioersted as in the German specification, and it is hoped that further improvements are in the realm of possibility.

It should be noted that the results reported here are of a very sketchy nature. A systematic study of the problems involved is now in progress at the Allegheny Ludlum Steel Corporation under Signal Corps sponsorship. It is hoped that this will result in the commercial production of silicon-iron materials designed for use in communication transformers in this country, and it is expected that the amount of nickel-iron alloys presently used for this purpose can be substantially reduced.

References

1. The Behavior of Iron and Steel Under the Operation of Feeble Magnetic Forces, **Lord Rayleigh.** *Philosophical Magazine*, London, England, vol. 24, 1887, pp. 225–45.

2. Ferromagnetic Constants for Weak Fields, **H. Jordan.** *Elektirsche Nachrichten-Technik*, Berlin, Germany, vol. 1, 1924, pp. 7–29.

3. Magnetic Measurements at Low Flux Densities Using the Alternating Current Bridge, **V. E. Legg.** *Bell System Technical Journal*, New York, N. Y., vol. 16, 1936, pp. 39–62.

4. Soft Iron for Communication Transformer Laminations (in German). *DIN E 41301*, German Standard Specifications, Dinorm, Berlin, Germany, Aug. 1943.

5. Magnetic After-Effects at Higher Inductions, **J. L. Snoek.** *Physica*, The Hague, Netherlands, vol. 6, 1939, pp. 797–805.

6. Iron-Silicon Alloys with High Initial Permeability Produced by Special Metallurgical Treatment, **F. Pawlek.** *Archiv fuer das Eisenhuettenwesen*, Duesseldorf, Germany, vol. 16, 1943, pp. 363–66.

◆

Discussion

D. C. Dieterly (Armco Steel Corporation, Middletown, Ohio): Dr. Both is to be congratulated for this fine paper calling attention to the characteristics of his experimentally processed material at low field strengths. This type of development has received little attention in this country, partly because of the lack of a clear demand for the product, but primarily because there has been such a large unsatisfied demand for grain-oriented silicon-iron alloys processed for low losses and high permeability at high flux densities.

Because our regular commercial grades of grain-oriented silicon-iron alloys have been aimed primarily at high-density applications, our published data were obtained for the sort of operating conditions where most of the tonnage is used. Consequently, Dr. Both did not have either our best available data at very low inductions, nor was he aware of the large amount of data on experimental variations from our normal product. Therefore, we should like to comment on a few factors which may not be generally known or which are apt to cause some misleading conclusions.

Armco Steel Corporation has been a large-scale producer of grain-oriented silicon-iron alloys for 15 years; and for the last 10 years has been producing highly oriented material as thin as 0.001 inch. The performance characteristics of such materials, not only in the high-induction range but in the Rayleigh region as well, have received considerable study.

Our commercial grades of thin, grain-oriented silicon-iron alloy have long been produced by the same process as Dr. Both's experimental material except that the final annealing is normally chosen to produce material suitable for the usual high-induction applications. However, materials intended for special annealing treatments have been available when needed to meet special requirements, and when the demand warranted a new grade. These processes have been patented[1] and licensed for use by others.

Of the three grades of Armco materials listed in Table I, only 2-mil Armco TRAN-COR T-O has been given any processing especially calculated to give a lower Rayleigh constant than would normally result from the processing given for good high-induction properties. Permeability and hysteresis data on 2-mil Armco TRAN-COR T-O in the range from 1 millioersted upward have been published,[2] calling particular attention to initial permeability greater than 1,500 and very low hysteresis loss at low inductions.

A few of the factors involved in the low-induction testing of cores which can have considerable effect on values of so-called initial permeability obtained on materials at these low magnetizing forces deserve some mention here.

Our experience leads us to conclude that there is no unique value of true initial permeability that can be assigned to a given specimen of these materials. The apparent value of initial permeability varies with frequency by an amount significantly larger than test error. Drift of permeability with time after demagnetization, while not nearly so large in these materials as in some others, is still of significant magnitude in part of the test range. These factors make it risky to specify a value of permeability at extremely low field strengths unless the test procedure is also specified in such detail that comparisons can be made on the basis of values obtained by identical procedures.

Another factor of considerable importance in obtaining measurements of such minute exciting currents while using electronic instruments such as oscillators, vacuum-tube voltmeters and oscillographs, is the likelihood that minute leakage currents circulating back through the power supplies of these instruments to a common ground will affect the readings to an unknown degree. Such troubles can often go undetected unless they lead to obviously improbable results. In testing at the lowest levels of magnetizing force mentioned by Dr. Both, we find it necessary to employ a special coupling transformer between the oscillator and test specimen to insure an exceptional degree of isolation from such leakage currents. Then, too, the use of battery-operated vacuum-tube voltmeters, together with the judicious selection of a common grounding point between these electronic instruments, helps in getting sound data by minimizing chances of leakage and pickup. These factors, too, make it important that the effect of circuit connections, instrumentation, and procedures be investigated thoroughly and described in detail.

In Table I Dr. Both lists a value of 1,000 for the estimated permeability of 2-mil Armco TRAN-COR T-O at 0.01 millioersted based on extrapolations of our catalogue data which were obtained by a-c bridge tests on a demagnetized specimen after 10 hours' drift. Our extrapolation from the a-c test data (taken down to 3 millioersteds) gives a value of 1,670 for the permeability of this specimen at 0.01 millioersted, when measured immediately upon demagnetization. Extrapolation of d-c tests made on the same specimen down to slightly less than 1 millioersted gives a value of 1,930 for the permeability at 0.01 millioersted. These test values correspond to a permeability gradient of 1.1 per cent per millioersted for the 0.01-to-100 millioersted range by a-c tests, and only 0.80 per cent per millioersted by d-c tests. If we use the 1.1 value of gradient and an initial permeability value of 1,670 as being most comparable to Dr. Both's procedure, the correct location for the point representing 2-mil Armco TRAN-COR T-O in Fig. 5 paper is much more nearly in line with his experimental material of equal initial permeability level than his estimate of 2.6 for gradient and 1,000 for the initial permeability.

Although we feel that the 2-mil material comes closest to being a practical means of meeting the requirements for these specialized applications, our regular 1-mil Armco TRAN-COR T-O has superior constancy of permeability in the Rayleigh region. D-c test data on this grade give a permeability of 1,380 at 0.01 millioersteds by extrapolation from d-c tests as low as 1 millioersted, and a slope of 0.51 per cent per millioersted in the 0.01-to-100-millioersted range. This grade has the smallest Rayleigh constant of any of our regular materials.

Regardless of what test method we might choose in which to place greatest reliance at very low field strengths, we believe Dr. Both's use of what we call the a-c crest-current method of permeability measurement is quite acceptable for comparing cores of the same dimensions, the same basic material, and similar windings, to note the trend of differences. It is a method that rates the performance of cores in proper relationship to the factors generally considered to be of paramount importance in some communications components. However, we do feel that the use of slope criteria based on a test point at 10 microersteds puts too much emphasis on a point that can easily be of dubious accuracy. At such field strengths we observe many anomalies in the characteristics of material of this type which affect the test results to a marked degree, but are not too well understood. It is our feeling that a much more dependable and reproducible value for Dr. Both's slope criteria is obtainable over a range of perhaps 0.20 to 20 millioersteds, or perhaps even 1.0 to 20 millioersteds. Even though these tests may not seem to be made at sufficiently low field strengths, there is no reason to suspect that, when the slope is sufficiently low in the range of 1 to 20 millioersteds, it will not be proportionally low in the range 0.01 to 1.0 millioersted. We feel, too, that the extension of slope criteria up as high as 100 millioersteds may be unduly severe, unless the requirements of the application clearly indicate operation at such relatively high fields.

The foregoing discussion should not be construed as implying that we disagree with Dr. Both's experimental test results or with his procedures. On the contrary, we find that his data agree very well with our own on such experimental materials. We are heartily in accord with his suggestion that silicon-iron alloys, processed in the manner he describes, be considered for communications components where it is indicated that a relatively constant level of permeability over a wide range of low field strengths is more important than a high average level of permeability over that range. His presentation of the experimental data on these thin silicon-iron alloys is a valuable contribution to the literature. It is hoped that it will result in concrete efforts to determine how cores of these materials compare with other materials both from performance and cost standpoints.

Table IV. Correlated Data in Interpretation of δ

Material	Anneal	μ_0	λx10⁴	δ
35 Permalloy[1]	1,000 C	1660	30	5.0
45 Permalloy[1]	1,000	2550	5.4	1.4
2.4–78 Cr. Permalloy[1]	1,000	14600	6.4	9.3
Vac Trafoperm *25NI*		830	26	3.0
2 mil 3.25 Si Steel	600	870	16	1.4
2 mil 3.25 Si Steel	900	1330	56	7.5
5 mil 3.25 Si Steel	600	770	43	3.3
5 mil 3.25 Si Steel	900	1950	127	24.7

This discussion is undertaken primarily to explain the ideas and position of one producer of such alloys, so that it would not be construed from the dearth of published information that commercial producers of magnetic materials in this country have not given consideration to, and investigated the properties of, materials for these specialized types of applications. Workers in the industrial field are seldom able to report any appreciable amount of the extensive experimental work which is undertaken in connection with the development of commercial magnetic materials. It is fortunate that careful, experienced workers like Dr. Both can perform a service of this kind, and see to it that the possibilities in some of these seemingly marginal applications are not overlooked.

REFERENCES

1. PROCESS FOR DEVELOPING HIGH MAGNETIC PERMEABILITY AND LOW CORE LOSS IN VERY THIN SILICON STEEL, **M. F. Littmann**. United States Patent No. *2,473,150*, June 14, 1947.

2. THIN-GAUGE IRON-SILICON ALLOYS FOR HIGH-FREQUENCY APPLICATIONS, **D. C. Dieterly**. *Electrical Engineering*, vol. 68, April 1949, pp. 335–36.

V. E. Legg (Bell Telephone Laboratories, Murray Hill, N. J.): Dr. Both is to be congratulated for the encouragement he has given to the steel industry to extend the utility of silicon iron. This is important in view of the danger which a nickel shortage would present for magnetic materials such as those needed by the Signal Corps, and by the communications industries.

Silicon steel has had a broad and intensive development as a "work-horse" for the electric power industry, where its behavior at high inductions has been of greatest interest. At the same time, its performance at low inductions has been disappointing, because of two handicaps: the rapid drop in permeability as the magnetizing force drops, and serious instability of permeability at low inductions—instability caused by lapse of time, particularly after previous magnetization at inductions higher than normal.

The paper indicates encouraging progress in removing one of these handicaps, through a combination of especially hard rolling of silicon steel sheet, together with closely controlled final heat treatment. This process evidently yields material with permeability declining much more slowly with decreasing field than that of normally fabricated silicon steel. Whether the other handicap is removed remains for examination. We would inquire if the author has data indicating improved stability of permeability with age after previous magnetization. If such data are not available, they will surely be needed before the industry can be confident that his type of silicon steel can be used in communication equipment.

In connection with Dr. Both's display of data, it is urged that new material constants should be avoided if possible, in the interest of clarity in conveying information. A new permeability variation constant δ has been defined in the text, and, rather inconsistently, interpretations of the same symbol but differing by a factor of 10 have been used in the tables. To correlate these data with previous literature, we offer Table IV extracted for the first three materials from a 1937 study of magnetic losses[1] and computed for the rest from the present paper.

In Table IV the fundamental definition of $\delta = (\mu-\mu_0)/\mu_0 H$ has been used, with the data of the present paper covering the range of field strengths 0.01 to 100 millioersteds. Obviously δ is the fractional increase in permeability per unit of field strength. For appraisal of high-permeability magnetic materials we have found it preferable to observe the variation of permeability as a function of induction B. The material constant λ in Table IV is the fractional increase in permeability per unit flux density B. It is related to δ by the equation $\lambda = \delta/\mu_0$. This latter constant is especially useful in calculating permeability variation as a function of voltage across a coil.

REFERENCE

1 STUDY OF MAGNETIC LOSSES AT LOW FLUX DENSITIES IN PERMALLOY SHEET, **W. B. Ellwood, V. E. Legg**. *Journal of Applied Physics*, New York, N. Y., vol. 8, 1937, p. 351.

R. A. Chegwidden (Bell Telephone Laboratories, Murray Hill, N. J.): The paper is of particular interest since it describes a type of 3 per cent silicon-iron having magnetic qualities approaching those of the higher priced and sometimes difficult-to-get 45–50 per cent nickel-iron alloys.

A sample of specially prepared 0.001-inch-thick material, which Dr. Both sent us, was found to be the most stable silicon-iron we have tested. Stability in this sense refers to the stability of its permeability at very low-flux densities with respect to mechanical shock and with time after demagnetization as described by Mr. Legg.

The 0.004-inch commercial grade samples mentioned by Mr. Dieterly were also very good in this respect, although not quite as stable as the sample from Dr. Both. This may be, as Mr. Dieterly suggests, because 0.004-inch and not 0.001-inch material was used.

Some types of communication apparatus will work at very low power levels where the current in the windings also is extremely small. For such applications, it is customary to use nickel-iron alloys for the core material. The true initial permeability of the core material is an important design factor in this case. It was not found possible to use 0.014-inch thick silicon-iron in these applications because not only is magnetic stability a problem, but also the permeability tends to fall off so rapidly with decreasing flux density at very low flux densities that the indicated initial permeability is sometimes not very much greater than that of iron.

We are encouraged to learn from the tests of Dr. Both's material and the 0.004-inch commercial grade silicon-iron that in these thinner gauge materials substantial improvements have been made in magnetic stability and in the permeability at very low flux densities. Measurements of these materials using magnetizing forces of the order of 1.0 millioersted or less, however, still show the tendency of the permeability to fall at an increasing rate with decreasing magnetizing force. The true value of the initial permeability is thus uncertain. This behavior is not encountered in the use of the 45–50 per cent nickel-iron alloys which have practically constant permeability at these low magnetizing forces. The permeability of the silicon-iron alloys measured at magnetizing forces of a microoersted or less would be of interest.

Eberhard Both: Mr. Dieterly points out correctly that, while the Armco materials, 14-mil-thick TRAN-COR 3X-O and 4-mil-thick TRAN-COR T-O, are most useful for high flux density applications only, the TRAN-COR T-O materials 2 mils and 1 mil thick have interesting properties for low flux density applications as well, and that the characteristics of the light-gauge materials approach those of the author's experimental material.

A comparison of these characteristics based on new and revised data furnished by Mr. Dieterly is given in Fig. 6 and Table V, where the curves labeled 1, 2, and 4 represent TRAN-COR T-O material 2 mils thick; and curves 3 and 5, two of the author's specimens of the same thickness. Curve 1 shows the TRAN-COR T-O material in the demagnetized state, and curve 2* after a 10-hour drift subsequent to demagnetization; curve 4 is a TRAN-COR sample deliberately annealed at a temperature below that recommended for 100 per cent relief of mechanical stresses. Curves 3 and 5 are replotted from Fig. 3; they represent specimens annealed at 850 and 750 degrees centigrade respectively and they were selected for coming closest to the permeability values of the TRAN-COR T-O samples.

Table V contains the numerical values for these curves. To represent the slope in the initial field region, a column for $\delta_{0.01-20}$-values has been added to those for $\delta_{0.01-100}$ and δ_{20-100} used in the paper. Comparing curves 1 and 2 with curve 3, and curve 4 with curve 5, it can be seen that the $\delta_{0.01-20}$-values of the experimental material are only about one-half of those for the TRAN-COR T-O 2 mils thick. While this is still a considerable difference, it is much smaller than the author was led to believe from the data at his disposal prior to Mr. Dieterly's communications.

The relatively good low flux density properties of the TRAN-COR T-O materials 2 mils thick may appear to be at variance with the author's statement that materials with low power losses have low initial permeability and high δ- values. As has been shown, the permeability and δ-values of the TRAN-COR T-O material 2-mil are not the best possible characteristics; furthermore, the power losses of this material are about 1.5 to 2 times higher than those of TRAN-COR 3X-O 14-mils and TRAN-COR T-O 4 mils thick*. The characteristics of 2-mil TRAN-COR T-O may therefore be said to constitute a compromise of intermediate properties for high and low flux densities rather than a combination of optimum properties for both flux ranges.

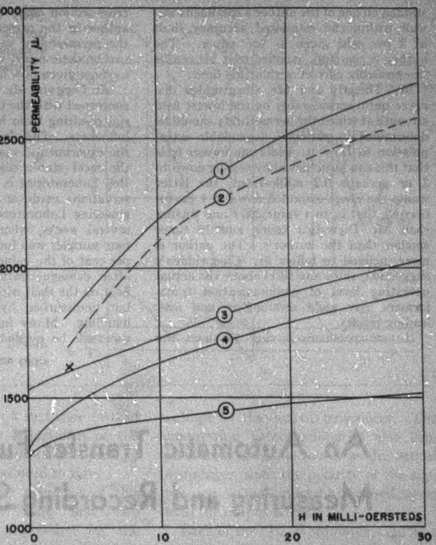


Fig. 6. Permeability of silicon-iron materials 2 mils thick:

a. Armco TRAN-COR T-O material, according to data furnished by D. C. Dieterly:
Curve 1—Demagnetized(no.5,371)
Curve 2—After 10 hours drift (no. 5,006, corrected)
Curve 4—Low temperature anneal (no. 5,372)

b. Experimental material, replotted from Fig. 3:
Curve 3: 850 degrees centigrade anneal
Curve 5: 750 degrees centigrade anneal

If optimum properties for communication transformers are required, it is necessary to accept a large increase in power losses, which are of no consequence for low flux density applications anyway. It is then possible to achieve characteristics equal to those shown by the author, in materials 4 and 14 mils thick, where the normal grain-oriented variety is so extremely poor, not in spite of, but as a consequence of its very excellence at high flux densities.

The author is aware of the pitfalls inherent in the testing procedures for establishing magnetic characteristics at extremely low flux levels. A complete description of these procedures is not possible here; however, it can be stated that: 1. The instruments used were calibrated individually for the frequency and sensitivity ranges in which they were used; 2. The complete circuit was calibrated by testing a linear inductor (Mo-Permalloy powder core) at the same levels of current and voltage employed in the silicon-iron core tests; and 3. The result of the test performed by Mr. Cheg-

Table V. Comparison of Characteristics Based on New and Revised Data

Curve	Material	Permeability at 0.01 Millioersteds	Permeability at 20 Millioersteds	Permeability at 100 Millioersteds	$\delta_{.01-20}$ per cent per Millioersted	$\delta_{.01-100}$ per cent per Millioersted	δ_{20-100} per cent per Millioersted
1	TRAN-COR T-O 2 mils. Demagnetized	1,670*	2,500*	3,500*	2.5	1.10	0.50
2	TRAN-COR T-O 2 mils. After 10 hours "drift"	1,500†	2,400	3,600	3.0	1.40	0.63
3	Experimental Material 2 mils. Anneal 850 degrees centigrade	1,520	1,900	2,950	1.3	0.94	0.69
4	TRAN-COR T-O 2 mils. Low temperature Anneal	1,300*	1,800*	2,790*	1.9	1.15*	0.69*
5	Experimental material 2 mils. Anneal 750 degrees centigrade	1,270	1,480	1,950	0.8	0.54	0.40

*Values furnished by Mr. Dieterly.

†Value estimated from corrected data on Armco catalogue specimen, furnished by Mr. Dieterly.

---

*Curve 2 is included here to correct an error which the author had copied from the Armco catalogue and which resulted in wrong values for the estimated initial permeability and δ used in Figs. 2 and 5 and in Table I. According to Mr. Dieterly, the permeability of 2 mil TRAN-COR T-O at $B = 5$ gausses is not 1,620 as shown in the Armco catalogue curve no. 5,006 (this point is marked by a cross in Fig. 6), but 1,695; consequently the estimate of the extrapolated initial permeability is revised from 1,000 to 1,500 and the $\delta_{0.01-100}$- value from 2.6 to 1.4 per cent per millioersted (see Table V).

*No way has yet been found to achieve the low power loss characteristics of 14-mil and 4-mil grain-oriented material in gauges of 2 mils or less; despite this handicap, however, these materials are extremely valuable for power applications at high frequencies, where eddy current losses are equal to or greater than the hysteresis losses.

widden on one of the author's specimens was well within the estimated accuracy limit of 2 per cent given in the paper. The author is therefore satisfied that his results are probably correct within this limit.

Mr. Dieterly and Mr. Chegwidden disagree quite considerably on the lowest field strength at which the permeability should be determined to permit a dependable extrapolation to $H=0$; while the former feels that this can be achieved by testing down to 1 or perhaps 0.2 millioersted, the latter wants the range extended down to 1 microoersted, that is, to a value 200 times smaller than Mr. Dieterly's figure, and 10 times smaller than the author's. The author is more inclined to follow Mr. Chegwidden's suggestion, since any tests above the actual operating level of communication transformers may yield meaningless and misleading results.

If test specimens having 10 times the cross-section area of those used by the author in the present study are available, the permeability values at 1 microoersted can be determined with the same accuracy as those given now for 10 microoersteds.

Mr. Legg and Mr. Chegwidden are mainly concerned with the instability of the material resulting from lapse of time and other influences. The latter's statement that the experimental material was found to be the most stable silicon-iron tested at the Bell Laboratories is in agreement with observations made at the Signal Corps Engineering Laboratories and extending over several weeks, where the permeability of four samples was found to remain within 2 per cent of the value obtained immediately after demagnetization, if the cores were kept on the shelf without a magnetic shield, but undisturbed by mechanical shock or handling. Many further tests of this nature will be needed in order to ascertain the limitations and possibilities of this material.

The author has used the δ-values for the variation of permeability rather than Mr. Legg's λ-values, because they are more familiar to him. While Mr. Legg is correct in stating that numerically the δ-values as expressed in per cent per millioersted are smaller by a factor of 10 than those which would follow from using the equation for δ as given in the paper, the values as given do represent the correct slope. In the German specification, the slope is expressed in per mille per millioersted, or $(\mu-\mu_0)/\mu$ per oersted, which results numerically in the δ-values as used by Mr. Legg. However, the author preferred to base his slope values on the percentual variation of permeability rather than on the variation expressed in per milles, because the latter unit appears to be hardly used in the technical literature of this country.

An Automatic Transfer Function Measuring and Recording System

R. J. EHRET
NONMEMBER AIEE

E. F. HOCHSCHILD
NONMEMBER AIEE

J. M. EMBREE
NONMEMBER AIEE

E. C. GROGAN
ASSOCIATE MEMBER AIEE

IN making frequency response tests of various devices, the phase and amplitude of the output must be measured when a signal having a known phase and amplitude is applied to the input. These data may be recorded in the form of a polar, or Nyquist, plot of amplitude with phase, frequency being varied as the parameter. If the input is set as the reference corresponding to a unit vector with its tip at the point (1,0) on the polar plot, then the locus of the vector representing the output of the device on the same polar plot will be the transfer function of the device throughout the range of frequencies being tested.

The method to be described eliminates some of the drawbacks of the "brute force" method of taking frequency response tests, which usually involves taking reams of oscillograms of the input and output wave forms and graphically measuring the phase and amplitude from the curves. A further difficulty lies in the need for having recorders with a dynamic response considerably better than that of the device under test, lest the recorder response give phase and amplitude errors. The system described here does not depend for its accuracy upon the dynamic characteristics of the recorder. A further advantage lies in the fact that the plot is made directly on a single sheet of paper, which simplifies the data reduction, handling, and storage problems considerably.

Operating Principles

Other authors[1] have described direct reading methods of obtaining frequency response tests whereby the output of the device to be tested is compared to a signal, the amplitude and phase of which may be adjusted to match the output exactly. An oscilloscope is used as the balance indicator. The advantage in the method described in this paper lies in the fact that the measuring and plotting is done by a recorder. Also, at very low frequencies, it is difficult to use an oscilloscope as a balance indicator.

Fig. 1 illustrates the general method of obtaining the transfer function. A sinusoidal test signal, $A \sin \omega t$, is applied to the device under test which has a transfer function, $ge^{j\theta}$ where g is the amplitude ratio and θ the phase shift. The output of the device will then be $gA \sin(\omega t+\theta)+C$, which can be represented by a vector having a magnitude gA making an angle θ with the input vector, plus a constant output term C. By suitable resolving means, the X and Y components of the output vector may be measured and used to position the two co-ordinates of a function plotter which traces the vector locus.

This method of resolving the X and Y components is accomplished by sampling the output wave at certain points in the cycle, these points being chosen so that they will be indicative of the X and Y components of the output vector. This sampling technique, has been used for measurements at power-line frequencies.[2]

This sampling is performed by means of switches synchronously driven from the signal generator shaft. The switches energize each axis at the proper intervals. The sampled points are the 0- and 180-degree points of the input wave for the Y co-ordinate, and 90 and 270 degrees for the X co-ordinate. Owing to the fact that the sine is negative in the 3rd and 4th quadrants, and the cosine is negative in the 2nd and 3rd quadrants, the output signal is reversed when the wave is sampled at these points. This reversal is indicated in the figure by the switches and the box labeled -1.

Fig. 2 shows a plot of the wave forms at the input and output of the device under test. The middle sine wave is the input, with the sampling points indicated. Directly above is shown the output wave form on the same time scale. The dotted portion shows the wave form as it would be without the periodic reversal; the

---

Paper 53-293, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 11, 1953; made available for printing April 30, 1953.

R. J. Ehret, E. F. Hochschild, and E. C. Grogan are all with the Minneapolis-Honeywell Regulator Company, Philadelphia, Pa. J. M. Embree was formerly with Minneapolis-Honeywell Regulator Company and is now with United Aircraft Corporation, Hartford, Conn.

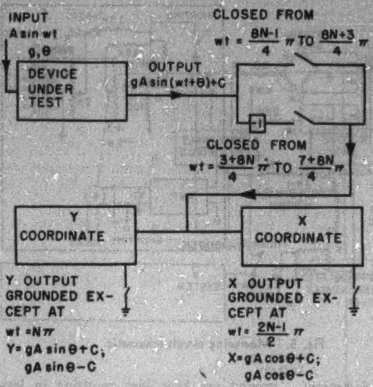


Fig. 1 (left). Block diagram showing principle of operation

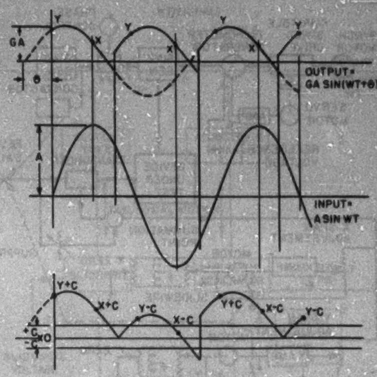


Fig. 2 (right). Wave form plots

solid portion shows the wave with reversals. The sampled points are indicated on the curves and designated as X for the points during which the X co-ordinate system is energized, and Y for the Y system. It is to be noted that for each axis two points are sampled per cycle.

It was mentioned previously that the output might contain a constant term C. This C might be any d-c component present in the output or a constant value of carrier in a carrier system. It might be caused by a zero shift present in the device or any static unbalance of the output transducer. To make use of the sampling technique for measuring the components of the vector, this constant term must be balanced out. This is accomplished by means of an adjustable suppression voltage in the measuring circuits.

The lower curve shows the effect of an error in this suppression adjustment. When the output of the device is reversed, it will also result in a reversal of the constant term C causing it to oscillate between $+C$ and $-C$ at the reversing frequency. Thus, the signals applied to the axes, which have this constant term added to them, will alternate between $X+C$ and $X-C$ for the X value, and $Y+C$ and $Y-C$ for the Y. This effect causes the point being plotted at a given frequency to appear as a rectangle instead of a point. Proper adjustment of the suppression control allows the area of this rectangle to be reduced to zero.

Table I shows the principle of this sampling technique. The first two columns show the expression for the output due to samples taken at 0 and 90 degrees without reversal. The vector sum of the X and Y components corresponding to the actual point plotted is $gA\epsilon^{j\theta}+\sqrt{2}C\epsilon^{j\pi/4}$. The last two columns show the sampled points at 180 and 270 degrees with the output reversed. It can be seen that the components result in the same vector except that the constant term now has the opposite sign.

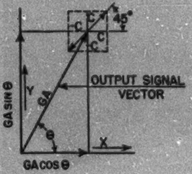
Fig. 3 shows a more complete block diagram of the system. The generator uses a closed-loop speed control servo to set and maintain the desired speed. Shaft rotation at this test speed is converted to pneumatic, positional, and electric sine waves by means of suitable transducers. This signal is then fed to the device under test.

The output of the device under test is converted to a 60-cycle modulated carrier signal by a suitable transducer, for example, a slide wire or a strain gauge. The reversing function described earlier is accomplished by reversing the supply voltage to the output transducer. The suppression potentiometer is also supplied from the same source as the output transducer, since the polarity of the suppression voltage must be reversed along with the output signal.

Each of the two co-ordinates includes a complete self-balancing system having an amplifier, balancing motor, and slide wire. Individual zero and span adjustments are provided for each axis, and serve the purpose of centering the plot and setting the proper scale factors. The output signal is connected to both systems; the sampling is performed by short-circuiting the balancing motors except during their sampling interval.

It is especially important to note that in the operation of the device the response of the recording system need not be fast, since the co-ordinates are required only to follow changes in the plotted point corresponding to the vector locus. This may be seen from a consideration of the fact that each co-ordinate is required only to balance to a single value at any one frequency, this point being the instantaneous value of the sine wave at the instant of sampling. Once a balance is reached by the system, no further movement of the co-ordinate drive system is required until the frequency is changed.

Table I. Development of the Polar Plot



	Output $= gA \sin(\omega t+\theta)+C$			
Polarity	Direct		Reverse	
Axis Sample instant	Y $\omega t=0$	X $\omega t=\pi/2$	Y $\omega t=\pi$	X $\omega t=3\pi/2$
Quadrature components transmitted to recorder	$gA\sin\theta+C$	$gA\sin(\pi/2+\theta)+C$ $=gA\cos\theta+C$	$-[gA\sin(\pi+\theta)+C]$ $=gA\sin\theta-C$	$-[gA\sin(3\pi/2+\theta)+C]$ $=gA\cos\theta-C$
Vector sum	$(gA\cos\theta+C)+j(gA\sin\theta+C)$ $=gA\epsilon^{j\theta}+C\sqrt{2}\epsilon^{j\pi/4}$		$(gA\cos\theta-C)+j(gA\sin\theta-C)$ $=gA\epsilon^{j\theta}-C\sqrt{2}\epsilon^{j\pi/4}$	

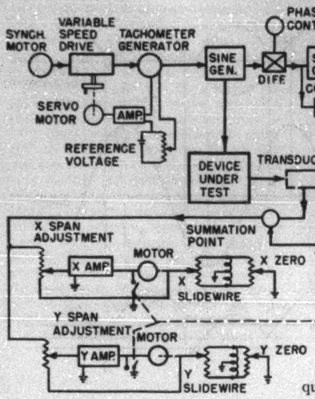

SYNCH. MOTOR
VARIABLE SPEED DRIVE
TACHOMETER GENERATOR
PHASE CONTROL
SINE GEN.
DIFF.
SINE GEN.
CONTACTOR
SERVO MOTOR
AMP.
REFERENCE VOLTAGE
DEVICE UNDER TEST
TRANSDUCER
REVERSING SWITCH
SUPPLY VOLTAGE
X SPAN ADJUSTMENT
SUMMATION POINT
X AMP.
MOTOR
X ZERO
X SLIDEWIRE
SUPPRESSION
Y SPAN ADJUSTMENT
MOTOR
Y AMP.
Y ZERO
Y SLIDEWIRE

Fig. 3. System block diagram

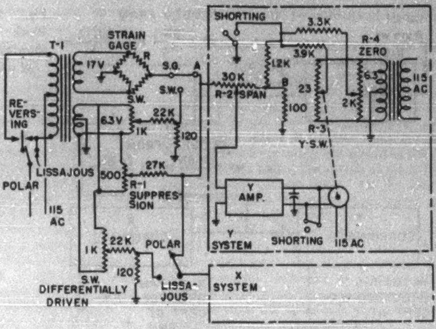

SHORTING
3.3K
R-4 ZERO
T-1
STRAIN GAGE
R
17V
S.G.
A
3.9K
12K
115 AC
30K
R-2 SPAN
B
23
6.3
S.W.
S.W.
RE-VERS-ING
6.3V
22K
1K
100
2K
R-3
Y-S.W.
120
LISSAJOUS
27K
500
R-1 SUPPRESSION
POLAR
Y AMP.
115 AC
1K
22K
POLAR
Y SYSTEM
SHORTING
115 AC
120
S.W. DIFFERENTIALLY DRIVEN
LISSA-JOUS
X SYSTEM

Fig. 5. Measuring circuit schematic

The rate of movement of the co-ordinates of a point is a function of the rate at which the frequency is varied. In practice, the frequency is varied manually in discrete steps, several cycles being allowed for a balance before the frequency is again changed.

The signal sampling interval is determined by the length of time that the short-circuiting switches remain open, and is influenced by two conflicting factors: the length of time required for motor drive and balancing, and the amount the instantaneous signal will vary during the sampled interval. It was found that about 6 degrees was a good compromise. For lower frequencies, a shorter interval can be used, since the sampled interval is then sufficiently long to obtain a balance in 1 or 2 cycles of the signal frequency. For higher signal frequencies, a somewhat longer period might be allowed, on the basis that the inertia of the motor will prevent any appreciable movement during 1 cycle.

Fig. 4 shows the entire system. The signal generator is in the lower left-hand corner. To the right of it is the control box which contains the operating controls as well as the electric circuits. In the top right-hand corner is the recorder, and adjacent to it the precision potentiometer, the d-c output potential of which determines the generator frequency.

Fig. 5 shows a schematic of the measuring circuits. Only the *Y* axis is shown in the diagram, but the *X* axis is identical in all respects. A 60-cycle suppressed carrier is used for all of the bridge and measuring circuits as well as for the amplifiers.

The use of a 60-cycle signal has the advantage that sufficient power is available to energize many types of output transducers, some of which require up to 1 watt. By using transformers having the center taps grounded, the impedances to ground are kept low, resulting in less trouble from electrostatic pickup. The use of the 60-cycle system has the disadvantage that signals must be properly phased. It also precludes the use of simple resistance-capacitance lead and filtering networks.

The strain-gauge output transducer and slide-wire signal transducer are energized from *T*-1, the primary supply to which is reversed to accomplish the signal reversal. The sampling switches short-circuit the amplifier windings of the balancing motors except during the sampling instant. It was also found necessary to short-circuit the voltage output of the slide wire of the axis not being sampled in order to prevent interaction between the two axis.

The bridge circuits use the principle of adding voltages through the use of summing resistors. The 3.3*K* and 3.9*K* act as summing resistances for the voltages at the slider of *R*-4 and *R*-3 respectively. Thus the voltage appearing at point *B* is the resultant of the slide wire and zero adjuster positions. The voltage which appears at point *A* is the resultant of the strain-gauge unbalance, and the position of the suppression potentiometer *R*-1.

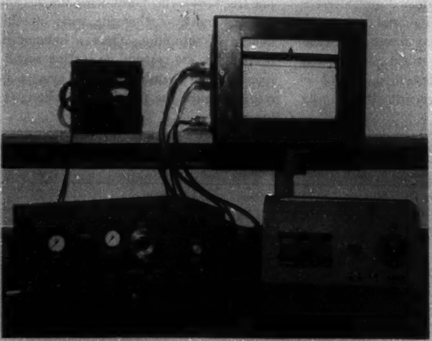


Fig. 4 (left). Entire system

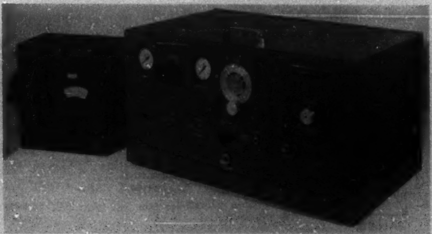


Fig. 6. Signal generator

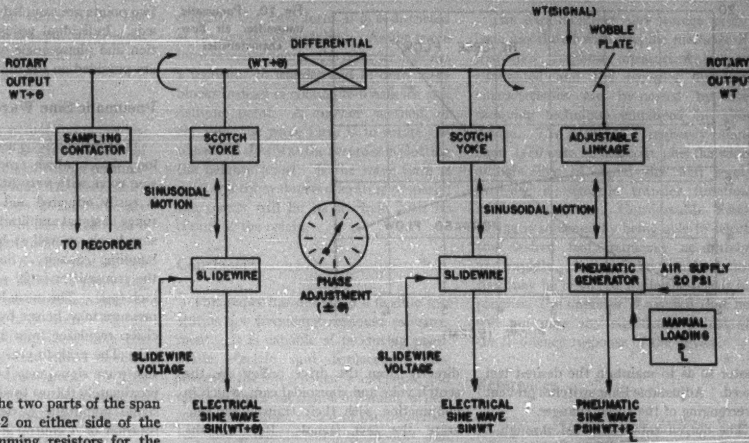

WT(SIGNAL)
WOBBLE PLATE
DIFFERENTIAL
(WT+θ)
ROTARY OUTPUT WT+θ
ROTARY OUTPUT WT
SAMPLING CONTACTOR
TO RECORDER
SCOTCH YOKE
SINUSOIDAL MOTION
SLIDEWIRE
SLIDEWIRE VOLTAGE
ELECTRICAL SINE WAVE SIN(WT+θ)
PHASE ADJUSTMENT (±θ)
SCOTCH YOKE
SINUSOIDAL MOTION
SLIDEWIRE
SLIDEWIRE VOLTAGE
ELECTRICAL SINE WAVE SINWT
ADJUSTABLE LINKAGE
PNEUMATIC GENERATOR
AIR SUPPLY 20 PSI
MANUAL LOADING P_L
PNEUMATIC SINE WAVE PSINWT+P_L

Fig. 7. Signal generator block diagram

In like manner, the two parts of the span potentiometer *R*-2 on either side of the slider act as summing resistors for the voltages at *A* and *B*.

For Lissajous operation, the two switches are thrown from "polar" to the "Lissajous." This connects the *X* system to the differentially driven reference slide wire, and cuts out the reversing switch in the primary of transformer *T*-1.

Signal Generator

The generator covers a frequency range from 1/30 cycle to 720 cycles per minute. It is capable of generating a wide variety of signals having good wave form and sufficient power output for most purposes. The desired frequency is set by a calibrated potentiometer and maintained by a speed servo. In addition to the test signals, reference and timing signals are provided. All controls and adjustments may be made from the front panel. Fig. 6 shows the generator and speed reference potentiometer.

Fig. 7 shows the block diagram of the generator. The main drive shaft rotating at the signal frequency is used to actuate all sinusoidal generating components. A wobble-plate sinusoidal cam activates the pneumatic generator through an adjustable linkage, as well as providing a motion sine wave suitable for testing low input-force devices such as pneumatic transmitters. Two scotch-yoke mechanisms, with a calibrated mechanical differential interposed between them, move the sliders of two linear slide wires. Also driven from the main shaft is the contactor which performs the quadrature sampling for polar operation.

For compactness the generator is arranged in two levels. The lower deck, Fig. 8, contains the main drive motor and speed control components. A 3,600-rpm synchronous motor drives a speed changer through a rubber-and-fabric timing belt. This stable speed changer consists of two disk-type variable speed drives in cascade, separated by a 6-to-1 gear reduction and further reduced at the output by a ratio of 12 to 1. A d-c tachometer generator develops a voltage proportional to the output speed, which is compared to a speed reference voltage from the precision portable potentiometer containing a standard cell. Any error voltage is applied to a d-c chopper-type amplifier to drive the speed change

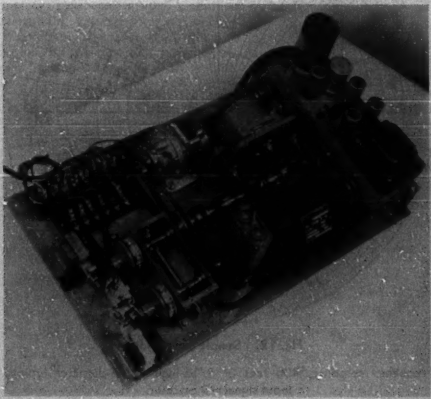


Fig. 8 (left). Lower deck

Fig. 9. Upper deck

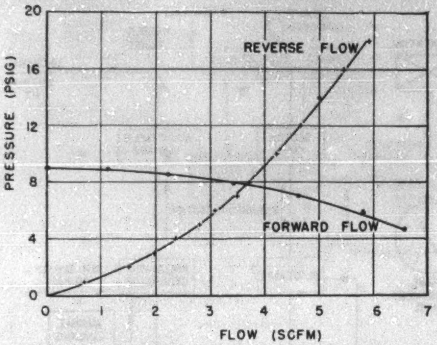


Fig. 10. Pneumatic transmitter air flow characteristics

motor so as to maintain the desired test speed. Adjustable limit switches prevent overranging of the speed changer.

The output rotation is fed through a gear box which gives a straight-through drive and also a 100-to-1 reduction. Either speed range may be selected by means of two solenoid-operated friction clutches controlled from the front panel. The driven members of the clutches are connected to the upper deck by means of a single timing belt. Extensive use in the drive unit of toothed timing belts eliminates many of the alignment and noise problems associated with gear trains.

The upper deck, Fig. 9, contains all of the signal-producing equipment. Driven directly from the drive pulley are the scotch yoke and sinusoidal cam, which in conjunction with their transducers generate the test signals. Between the generator elements and the reference and switching components is the differential gear which varies the relative phase. The reference and signal slide wires, which are standard Honeywell-Brown parts, are available in several resistances and are mounted so that they may be replaced quickly and easily. Also driven after the differential gear are the polarity reversing switches and the contactor unit. The contactor unit consists of four normally closed contacts which are opened periodically by a set of single-lobe cams driven at twice the test speed, since two points are sampled per cycle for each axis. Individual adjustments for duration and phase angle of contact opening are provided on the unit.

Pneumatic Sine Wave Generator

The pneumatic generator is a closed-loop device which produces an accurate sine wave with very little distortion. It is easily adjusted and provides a large range of signal amplitudes and mean pressure levels as well as having a high air-handling capacity. Supply pressure to the transmitter is 20 pounds per square inch (psi). The mean level of the output pressure may be set by means of a precision regulator from 1 psi to about 19 psi. The peak-to-peak amplitude of the sine wave signal may be varied from approximately 0.1 psi to about 15 psi. The signal amplitude is determined by the rotation of a control knob on the front panel of the drive unit. An adjustable linkage is also provided on the transmitter for setting the range of amplitudes.

Fig. 10 shows the air-handling capacity of the transmitter in the forward and reverse flow direction, which determines the pneumatic load that may be supplied from this generator. Frequency response tests have shown that the transmitter is flat to 300 cycles per minute (cpm) with a peak-to-peak amplitude of 1.2 psi at a mean level of 9 psi. These data were obtained when the transmitter was dead-ended into a strain-gauge pressure

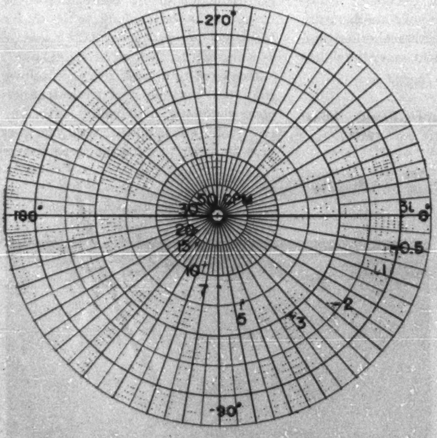


Fig. 11. Sample polar plot

Frequency response, 400 feet of 3/16-inch inside-diameter tubing plus 100-inch³ volume. Input signal = 9 psi ± 0.6

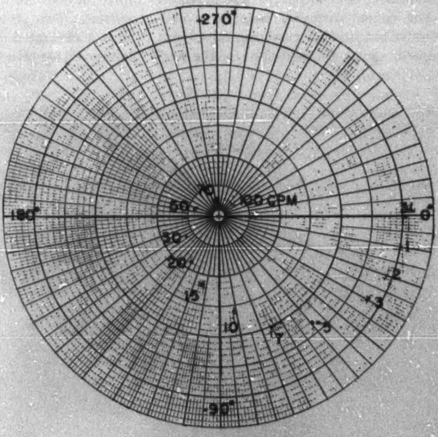


Fig. 12. Sample polar plot

Frequency response, 400 feet of 3/16-inch inside-diameter tubing. Input signal = 9 psi ± 0.6

transducer. For the same signal the transmitter is flat to 80 cpm when pipe-coupled into a 100-inch^3 volume.

Test Results

Fig. 11 shows an actual frequency response plot taken on a length of 400 feet of 3/16-inch inside-diameter tubing terminated in a 100-inch^3 volume. The frequencies corresponding to points on the curve are labelled and they cover the range from 0 to 50 cpm. The input to the tubing at all frequencies is shown on the plot as the point 3*i*. The plot shows a phase lag of 230 degrees and an amplitude ratio of 0.03 at a frequency of 50 cpm. Fig. 12 shows a similar test run using only the 400 feet of tubing dead-ended directly into the pressure transducer. In this case the plot indicates a phase shift of 200 degrees and an amplitude ratio of 0.13 at a frequency of 50 cpm.

The points indicated by 70 and 100 cpm on this plot illustrate the type of plot which results if the measured signal is not a good sine wave. In this respect the system provides a simple means of self-checking the wave form, if it is distorted at the instant of sampling. As the wave form becomes poorer, the recorder exhibits a tendency to wander and to draw large blocks instead of coming to a balance at a definite point. A further method of checking the wave form is to rotate the phase dial slowly as the recorder is plotting the balance point. If the wave form is good and the system is operating properly, the points will lie on a circle with its center at the origin.

Conclusion

This paper has described a flexible test station for frequency response measurement. It is capable of producing pneumatic, electric, and displacement sine waves in the low-frequency spectrum, and offers a quick means of obtaining a permanent, compact record of the frequency response characteristics of devices and systems. It has not only proved its usefulness in the laboratory for device and analogue testing but has also demonstrated its suitability and ruggedness in the field-testing of processes.

The data-handling and storage problems associated with graphic methods of frequency response measurements are eliminated and the recorder dynamic characteristics are by-passed by the sampling technique employed by this device. The 20,000-to-1 frequency range from 1/30 cpm to 720 cpm, plus the wide adjustability of signal size and level, provides a versatile transfer function measuring system. Considerable time-saving is achieved by being able to select the desired test frequency accurately and rapidly. Because the transfer function may be observed while the test is in progress, the operator is assured that he will not miss any significant points of the frequency response of the device.

References

1. Modern Servomechanism Testers, G. A. Korn, T. M. Korn. *Electrical Engineering*, vol. 69, no. 9, Sept. 1950, pp. 814–16.

2. Polar Vector Indicator, E. A. Walker, A. H. Waynick, P. G. Sulzer. *AIEE Transactions*, vol. 68, pt. I, 1949, pp. 154–59.

3. A Portable Instrument for Process Control Analysis, A. R. Aikman. *Instrument Practice*, London, England, vol. 5, no. 7, May 1951.

◆

Discussion

Page S. Buckley (E. I. du Pont de Nemours and Company, Orange, Texas): Although the determination of transfer-function data sometimes appears simple and straightforward in theory, in practice it is apt to be tedious and time consuming. This is because most of us must use such test equipment as oscilloscopes or oscillographs, laboriously pick off phase and amplitude, and plot it up on suitable graph paper. Any apparatus to mechanize or replace this procedure is welcome, and the authors have successfully developed such equipment.

This system is fairly elaborate and more expensive than conventional equipment for determining transfer-function data. This will probably be more than offset, in my opinion, by the time and labor saving features for three categories of users:

1. Research groups which have enough dynamic testing to provide a good load factor for the equipment.
2. Instrument or control equipment manufacturers who may wish to use dynamic tests for quality-control purposes.
3. Instrument departments in such manufacturing plants as chemical plants or oil refineries, where instruments and control devices get periodic overhauls and tests.

The design of this equipment appears to be basically sound and I have only two comments about it:

1. It has been our experience that there are several commercial variable-speed devices with adequate speed regulation, whose use would eliminate the need for a tachometer, d-c amplifier, and precision portable potentiometer.
2. The work we have done on dynamic testing of pneumatic instruments indicates the need for a transducer with a wider bandwidth and higher air capacity. Specifically, this reviewer suggests 0 to 10 cycles per second for loads of 10 standard cubic feet per minute or higher. This is needed to test pneumatically operated control valves, such as are used in the oil and chemical industries.

R. J. Ehret, E. F. Hochschild, J. M. Embree, and E. C. Grogan: We wish to thank Mr. Buckley for his thoughtful comments, which clearly point out the advantages of this method of taking transfer function measurements.

In reply to the two points which were raised by Mr. Buckley, we offer the following comments:

1. A closed-loop speed control system was chosen, not because the regulation was a problem, but rather because we desired an accurate and fast means of setting frequency. Furthermore, it is not necessary to depend upon any speed-changer dial calibration.
2. We agree with Mr. Buckley that for certain tests the pneumatic power requirements on a sine-wave generator are very high. For such tests, auxiliary equipment in the form of power amplifiers has been used. Even the suggested 10 cycles per second is not high enough to perform all testing; therefore other test equipment is used in the high-frequency spectrum.

Transmission Design of Intertoll Telephone Trunks

H. R. HUNTLEY
FELLOW AIEE

FOUR papers have already discussed such features of the nation-wide toll switching plan as the fundamental plant layout, numbering plan, toll switching, and automatic accounting equipments.[1–4] The present paper is intended to round out this coverage of the plan with a further discussion of the transmission features.

The Problem

In the new nation-wide toll switching plan using switching machines the layouts of toll circuits as well as the routings of traffic will be quite different from that of the earlier plans, which were based on manual switching. Individual calls can be switched so fast and so cheaply that switching is no longer a limiting factor, and circuits can be laid out and used in such a way as to obtain maximum economy with few, if any limitations from the switching standpoint.

An example of these changes is given in Fig. 1(A), which shows the circuit groups which would be used to handle a given (assumed) flow of traffic on a manual basis and in Fig. 1(B) the groups which would be used to handle the same traffic on a dial basis. In (A) there are 44 different circuit groups and in (B) there are 26 circuit groups. More specific ideas regarding the effects of these differences can be obtained by considering how calls between specific centers (for example, in Fig. 1, *A*1 to *B*1) would be routed in the two plans.

From the transmission standpoint the principal impact of the new plan is that the situation will be changed from one in which as much of the traffic as practicable was handled over direct circuits with a minimum of switched traffic (circuits in tandem) to one in which two or more (up to a maximum of eight) circuits will be used in tandem on many calls and in which different numbers and make-ups of circuits may be encountered on successive calls between the same two telephones, as a result of the alternate routings employed with machine switching. This means that the losses of circuits must be low to provide adequate transmission on all calls and to avoid large differences in transmission on successive calls between the same two places.

The ideal method in such a situation would be to operate all circuits at zero loss, since this would make the results independent of the number of circuits in tandem. However, the distances involved in the Bell System are so great that the propagation times, which affect echo, and the crosstalk between circuits require that even carrier circuits be operated at finite losses. Also, the plan must accommodate many voice frequency circuits on which the noise and singing conditions, as well as echo and crosstalk, may be more severe than on carrier circuits. The practical plan, therefore, is to:

1. Operate every circuit at the lowest loss practicable considering its length and the type of facilities used.
2. Assign circuits with different transmission capabilities in accordance with the parts they have to play in the operation of the over-all network.

The principal problem is to determine how low circuit losses can be made without getting into trouble as a result of one or more of the limitations mentioned above. This problem is complicated by the fact that the effects of these limitations are not directly proportional to circuit length or to the number of circuits in tandem. For example, if circuit *A* can be operated by itself at a loss of *X* decibels and circuit *B* can be operated by itself at a loss of *Y* decibels, the loss permissible when circuits *A* and *B* are switched together is less than $X+Y$. Ideally, therefore, each circuit should have a different loss in each different connection in which it is used. However, this is not practicable and a compromise must be adopted. This compromise provides that in some connections a particular circuit will operate at its lowest practical loss while in other connections higher losses will be employed to give over-all figures that will be adequate from the standpoint of echo, crosstalk, etc. The general procedure is as follows:

1. When a toll circuit is switched to another toll circuit at both ends work it at a loss which is called via net loss (VNL).
2. When the circuit is switched to another at one end only (the other end being at the point of origin or destination of the call) work it at a loss higher than VNL by an amount which we shall call *S*, *S* being a generic term derived from the fact that it may be associated with switching pads, usually called *S* pads.
3. When the circuit is used by itself (that is, the origin and destination of the calls are at the ends of the circuit) increase its loss by *S* again, that is, work it at VNL plus 2*S*. This is known as terminal net loss (TNL).

Via net loss is, of course, to be the lowest loss practicable referred to, and the next step is to establish methods of deriving VNL and of selecting the best value for *S*.

Since it would be a very complicated process to work simultaneously with all four of the limiting factors mentioned in the foregoing (echo, crosstalk, singing, and noise), the practical approach has been to select one of them as the basis of design and then check the results against the other three, modifying the final solution as necessary so that all four are kept under control. Since long experience indicates that echo is likely to be the most difficult and complex factor to control, it has been used as the starting point in the solution of the problem. As will be evident later, there are a large number of solutions possible from the echo standpoint, and the one which has been selected has been affected to a considerable extent by the other factors.

The next part of this paper is, therefore, devoted to an analysis of circuit design from the echo standpoint.

Determining Lowest Practicable Circuit Loss from Echo Standpoint

The over-all objective is to have practically no cases in which objectionable echo will be observed by customers.

If circuits could be precisely adjusted to the requirements in each different connection the probability of echo would be the same on all connections and the computations would have been carried out on the basis of a very small probability, say, 1 in 10,000. However, losses can be

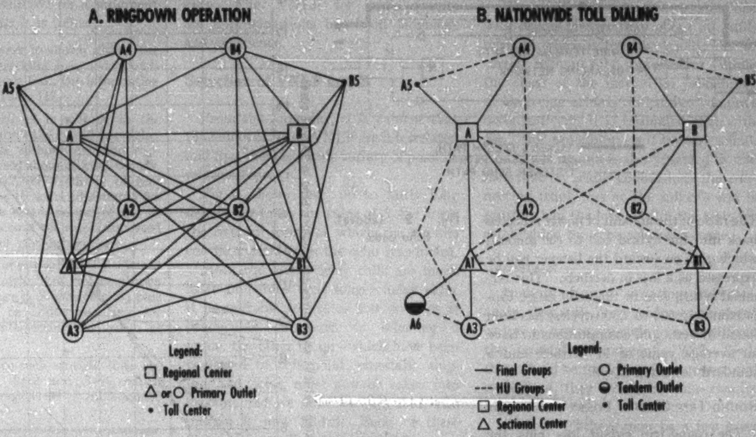
---

Paper 53-227, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 18, 1953; made available for printing April 22, 1953.

H. R. Huntley is with the American Telephone and Telegraph Company, New York, N. Y.

Many persons took an active part in the preparation or in the development of the background data for this paper. The author wishes to mention, particularly, L. L. Bouton, who did much of the basic work on the mathematical concepts involved, on the simplification of these concepts for practical application, and on the re-evaluation of data required in these applications.

Fig. 1. Typical intertoll trunk networks



changed only in discrete steps S so that in a very large proportion of cases the losses will be higher than theoretically necessary. Hence it seems sensible to compute the theoretical losses on the more liberal basis of 1 in 100, relying on the excess loss in most connections to reduce the over-all probability to the very small value desired.

The echo problem with which we are concerned is illustrated in Figure 2. As shown there, part of the speech power which is being transmitted to the listener leaks across the hybrid, or 4-wire terminating set, at the listener's end and returns to the talker. This is known as talker echo. Actually, of course, some of the echo which returns to the talker can leak across the hybrid there and go back to the listener. This is known as listener echo. However, with modern plant listener echo will not be important if talker echo is adequately controlled.

Considering the effect of echo on the talker, if the elapsed time before the echo gets back to him is very short, it is just like hearing his own voice through the sidetone in his own set, and unless it is very loud he does not notice it. If, on the other hand, the elapsed time is long, it sounds to him very much like the familiar acoustical echoes arising from physical obstacles. In extreme cases he may get the impression that the distant party is trying to interrupt him.

The over-all effect of echo then depends on:

1. How loud the echo is, which in turn depends on how much loss there is in the echo path.
2. How long it is delayed before it gets back to the talker.
3. How easily the talker is annoyed by it (his tolerance to echo).

The factors involved are discussed in more detail in the following. For simplification, 4-wire circuits and 4-wire switching are assumed; 2-wire circuits and 2-wire switching are treated as variations and are discussed later.

TOLERANCE TO ECHO

People vary greatly in their reaction to echo of given magnitude and delay, and it is therefore necessary to treat them statistically, using the usual two statistical terms to describe a mass of data, that is, average and standard deviation.

Average means simply the familiar algebraic average of the data. No talker is average but it is possible to obtain the average of a lot of talkers. Standard deviation means a number which tells in general terms how the individuals spread out on both sides of the average. Usually 30 to 35 per cent will be between the average and 1 standard deviation below the average; 30 to 35 per cent will be between the average and 1 standard deviation above it. At least 45 per cent will be between the average and 2 standard deviations below it and at least 45 per cent will be between the average and 2 standard deviations above it. Very few, if any, will be outside of ±3 standard deviations.

Judgment tests under controlled conditions and with a number of observers (talkers), under conditions simulating connections to subscribers near the toll office, have given the following basic data on these effects: (These data are slightly different from some published earlier because of recent re-evaluations. Further studies are now under way and may indicate some further changes.)

An analysis of all the test data indicated that the observer judgments conformed fairly well with a normal law curve having a standard deviation (D_0) of 2.5. This means, for example, taking the 40-millisecond delay condition, that while on the average a 17.7-decibel loss was required in the echo path to make the echo just tolerable, some 30 to 35 per cent of observers could tolerate 2.5 decibels less loss. Another 30 to 35 per cent were sensitive enough to need 2.5 decibels more loss for satisfactory echo condition. Practically no observer was so sensitive as to require $3 \times 2.5 = 7.5$ decibels more than 17.7 decibels, and practically none was so tolerant that he could permit 7.5 decibels less.

TERMINAL RETURN LOSS

As shown in Fig. 2, with 4-wire switching and 4-wire-type circuits, the only source of echo is lack of perfect balance between the balancing network of the 4-wire terminating set (hybrid coil or its equivalent) and the trunk, loop, and subscriber station connected at the customer side of the set at the distant terminal.

Table I. Results of Judgment Tests

Round-Trip Delay, Milliseconds	Loss in Echo Path Just Satisfactory to Average Observer, Decibels
0	1.4
20	11.1
40	17.7
60	22.7
80	27.2
100	30.9

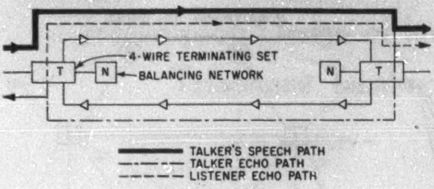


Fig. 2 (above). Echo paths

The ratio of the amount of power reflected back into the hybrid coil to the amount which goes on toward the listener can be expressed as a loss in decibels. This terminal return loss in the echo range (approximately 500 to 2,500 cycles) has been found by tests and computations to have an average value of 11 decibels and a standard deviation $D_t = 3$.

Round-Trip Circuit Loss

The round-trip circuit loss plus the terminal return loss is the total loss in the echo path. The round-trip circuit loss, that is, the over-all loss which the intertoll trunk (or trunks) inserts in the echo path, is the sum of the losses in the east-to-west and west-to-east directions. If the circuit regulation were perfect, this loss would simply be twice the nominal one-way loss of the trunks, which is the value sought.

However, regulation is not perfect, and to determine what the nominal loss should be, the deviations from it which are certain to occur in practice must be taken into account. A considerable amount of experience indicates that these deviations can be treated statistically and, considering some improvement in maintenance methods and procedures and a wider use of carrier systems with improved regulation, a standard deviation of $D_v = 2$ decibels for round-trip losses seems a not unreasonable assumption for the next few years.

Relationship Between Working Echo Net Loss and Round-Trip Delay

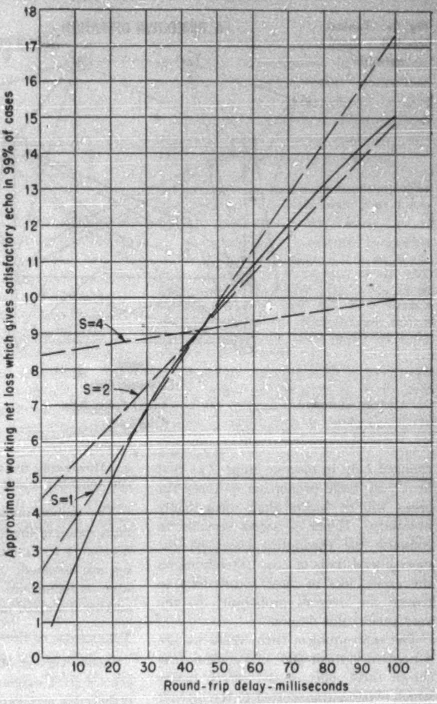
All of the data needed are now assembled to solve the problem, which, as stated before, is to find what VNL to assign to a circuit of given length and on a given type of facility.

The first step mathematically is to combine the three statistical distributions discussed, that is, tolerance to echo, terminal return loss, and the variations in round-trip circuit loss.

The combined standard deviation (D_c) of the three sets of distributions is the square root of the sum of the squares of the standard deviations of the individual distributions. The first two distributions are independent of the number of links N (assuming 4-wire switching) but the distribution of circuit loss variations is a function of the number of links. The mathematical expression for the combination of these three standard deviations is

$$D_c^2 = D_e^2 + D_t^2 + ND_v^2 = 2.5^2 + 3^2 + N2^2$$

Fig. 3 (right). Approximate relationships between round-trip delay and permissible working 1-way loss for an intertoll trunk from echo standpoint for 4-wire circuits and 4-wire switching



From this equation Table II can be constructed.

In line with the principles stated at the outset, the mathematics will be worked out on the basis that 99 calls out of 100 will be free from echo, then margins will be added. The mathematics are as follows:

1. To meet 99 per cent of the cases, 2.33 standard deviations must be used, or average round-trip loss = average echo tolerance—average return loss + 2.33 standard deviations.
2. The average 1-way loss is the loss to be assigned and is one half the average round-trip loss.

Permissible average losses with several different numbers of links, based on this equation, are given in Table III. These data are for 4-wire circuits and 4-wire switching. It will be noted that for a given total round-trip delay the necessary increase in over-all loss for increasing numbers of links varies somewhat for different conditions but for the more severe cases it is about 0.4 decibel per link.

Because, as stated at the outset, the relationship between round-trip delay and permissible circuit loss is not linear, Table III cannot be used directly in selecting the working net loss of a circuit to be used in switched connections. An example will make this clear:

1. From Table III, the permissible loss of a circuit with 20 milliseconds round-trip delay is 5.0 decibels.
2. If this were used as the basis for designing the circuit, the loss of four such circuits in tandem would be 20.0 decibels whereas the table shows that four links with a total round-trip delay of 80 milliseconds could be operated at 14.6 decibels.

The problem then is to find the best method of determining VNL and *S*. As indicated earlier, this problem has a wide variety of solutions, among which the best can be selected on a judgment basis. The process is as follows:

1. In Fig. 3 a curve is shown giving the relation between working loss and round-trip delay for a single link, the information being taken from Table III.

2. With the plant as it will be in the reasonably near future the round-trip delay on any connection without an echo suppressor will not exceed about 45 milliseconds. This figure is based on a survey of geographical lengths, with some adjustment for the expected more extensive use of carrier and taking into account the rules (discussed later) for the use of echo suppressors.

3. Then starting at any arbitrarily selected value of S, a straight line can be drawn from $2S$ (since there is S at each end) plus 0.4 (required to be added per link for variations) and intersecting the curve at 45 milliseconds.

4. In Fig. 3, three such straight lines are drawn, for $S=1$, $S=2$ and $S=4$, which have slopes (in decibels per millisecond) about as follows:

S	Slope
1	0.15
2	0.10
4	0.016

5. From these data, equations for VNL and TNL for 4-wire circuits can be worked out in terms of round-trip delay d thus:

S	VNL	TNL = VNL + 2S
1	$0.15d+0.4$	$0.15d+2.4$
2	$0.10d+0.4$	$0.10d+4.4$
4	$0.016d+0.4$	$0.016d+8.4$

The slopes of the lines can be converted into factors called via net loss factors (VNLF) in terms of decibels per mile, by dividing twice the slope by the velocity of propagation in miles per millisecond. The product of VNLF and circuit length in miles plus 0.4 gives VNL.

As an example, for K carrier circuits with a velocity of propagation of 105,000 miles per second, the via net loss factor for $S=2$ would be $(2\times0.10)\div105=0.0019$.

For 2-wire circuits the local echo paths at the repeaters make it impracticable to establish a straightforward relationship between over-all delay and echo performance. The VNLF for such circuits are approximations based on judgment and experience.

Table II. Deviation in Relation to Number of Links

No. of Links	Standard Deviation, Decibels
1	4.4
2	4.8
3	5.2
4	5.6
5	5.9
6	6.3
7	6.6
8	6.9

Selection of VNLF and S

From the foregoing, it is evident that the values of S and VNLF are interrelated and that there is a wide variety of possible relationships.

Fig. 3 shows that, up to fairly long delays, the lower the value of S, the lower the over-all losses at which the circuits can be worked from the echo standpoint.

Since the lower delay calls are much more numerous than longer delay calls, it is desirable to use as low an S as is practicable. However, in selecting a value, the other factors which have been neglected to this point, crosstalk, singing, and noise, must now be taken into account and we must be sure that echo margin is now added. Each of these factors is now discussed separately.

SINGING

The more extensive use of carrier reduces the importance of singing because voice frequency circuits are becoming shorter, thus eliminating the difficult singing problems associated with multi-repeater-section 2-wire circuits. On the other hand, some of the conditions at circuit terminals may become more severe from the singing standpoint.

Studies indicate that over-all losses obtained with $S=2$ are adequate to care for singing under most conditions, but that if $S=1$ were adopted, singing would be more important. With $S=2$ the necessity for increasing circuit losses to avoid excessive danger of singing will probably be confined to a few open-wire circuits having large discrete irregularities.

NOISE

Noise is usually not a factor in the assignment of circuit losses. Carrier systems are so designed that under normal conditions the noise is low enough so that any desired loss can be used. If, in a specific case, noise in either carrier or voice frequency circuits is too high, the approach is to get rid of it by one or more of the means available.

ECHO MARGIN

Reference to Fig. 3 will indicate that for $S=2$ there is 2 decibels or more round-trip echo margin in all cases with round-trip delays less than the order of 30 or so milliseconds. With $S=1$ there is little margin and with $S=4$ there is excessive margin.

There will be very few connections with delays greater than about 30 milliseconds, which is roughly 1,500 miles of carrier, without echo suppressors. While the effect of the margin on probability of observing echo is difficult to compute quantitatively, it is estimated that with $S=2$ this probability will be very small. Additional margin is also provided by the fact that on many connections the connecting trunk loss at the talker's end is greater than the loss used in the tests for establishing the echo tolerance curve.

CROSSTALK

Analysis of many situations indicates, again, that with $S=2$, the losses are about as low as are practicable in general from the crosstalk standpoint. With $S=1$, they would be too low in many cases, and with $S=4$, they would be unnecessarily high.

With $S=2$, there may be a few cases where specific attention to crosstalk will be needed, particularly in open wire.

FINAL SELECTION

From the consideration of factors like the foregoing the value of 2 decibels has been selected for S on a judgment basis.

Provision of S

The loss S can be provided in any of the following ways as appropriate (see Fig. 4):

1. As a switchable loss pad in the intertoll trunks.

2. As a fixed loss pad in the toll connecting trunk.

3. As part of the conductor loss of toll connecting trunks. This can be done only if the structural return loss of the connecting trunks against the balancing network is reasonably good.

4. If there is to be no switching to other intertoll trunks or to connecting trunks with more than a 2-decibel loss, S may be provided simply by increasing the circuit loss by 2 decibels.

Table III. Permissible Average Losses in Relation to Round-Trip Delay

Total Round-Trip Delay, Milliseconds	Permissible Working Over-all 1-Way Loss, Decibels			
	1 Link	2 Links	4 Links	6 Links*
0	0.3	0.8	1.7	2.5
20	5.0	5.6	6.5	7.4
40	8.5	9.0	9.8	10.6
60	10.9	11.4	12.3	13.1
80	13.2	13.7	14.6	15.4
100	15.1	15.6	16.5	17.2

* With the switching arrangements which will be used, not more than six intertoll trunks will be used in tandem without an echo suppressor.

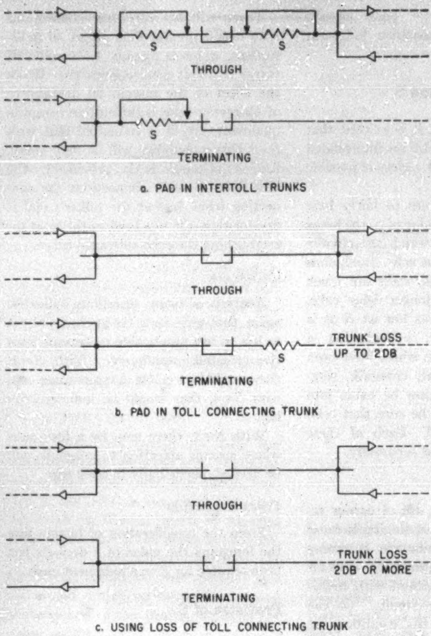
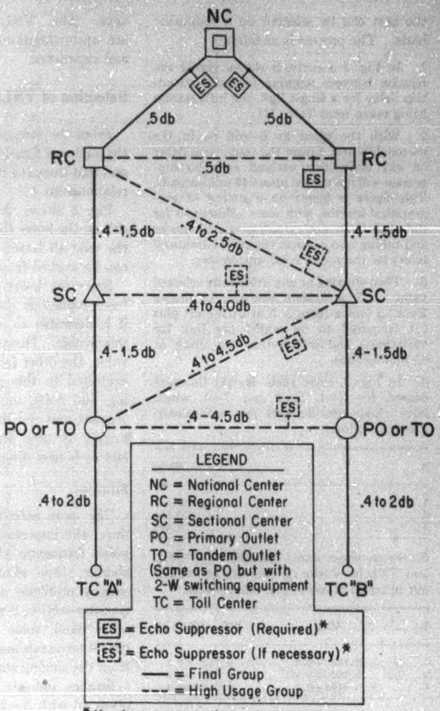

S
S
THROUGH
S
TERMINATING
a. PAD IN INTERTOLL TRUNKS
THROUGH
TRUNK LOSS
UP TO 2DB
S
TERMINATING
b. PAD IN TOLL CONNECTING TRUNK
THROUGH
TRUNK LOSS
2DB OR MORE
TERMINATING
c. USING LOSS OF TOLL CONNECTING TRUNK

Fig. 4 (above). Methods of providing S


NC
ES
ES
.5db
.5db
RC
.5db
ES
RC
.4-1.5db
.4 to 2.5db
ES
ES
.4-1.5db
SC
.4 to 4.0db
SC
.4-1.5db
.4 to 4.5db
ES
.4-1.5db
PO or TO
.4-4.5db
ES
PO or TO
.4 to 2db
.4 to 2db
TC "A"
TC "B"
LEGEND
NC = National Center
RC = Regional Center
SC = Sectional Center
PO = Primary Outlet
TO = Tandem Outlet
(Same as PO but with
2-W switching equipment
TC = Toll Center
ES = Echo Suppressor (Required)*
ES = Echo Suppressor (If necessary)*
—— = Final Group
- - - = High Usage Group

* If echo suppressor is used, the assigned loss is .5db unless a higher value is required for crosstalk.

Fig. 5. Intertoll routing pattern between two regions showing typical circuit groups

Via Net Loss Factors

Table IV lists typical VNLF's of Bell System intertoll trunk facilities for the condition $S=2$.

Typical losses at which circuits would be worked with $S=2$ and with the VNLF's tabulated in Table IV are given in Table V. The advantages of high velocity, 4-wire, circuits, carrier and radio, are obvious from these tables.

Echo Suppressors

Even if the intertoll trunk plant of the Bell System were all carrier, the length of some connections would be so great that some method of controlling echo other than simply increasing circuit loss would be desirable. Lower losses can be obtained on such connections through the use of the echo suppressor, an electronic device which under control of the talker's speech currents places a high loss in the return path at the right time to intercept the return echo currents.

Table IV. Typical Via Net Loss Factors

Facility	VNLF, Decibels per Mile 2-Wire Circuits	 4-Wire Circuits
19H-88-50	0.03	0.014
19H-44-25	0.02	0.010
O. W. voice	0.01	
O. W. carrier		0.0017
K or *N* carrier		0.0019
L carrier		0.0015
Radio		0.0014

Echo suppressors perform very well so long as not many circuits equipped with them are connected in tandem and there is not too much time delay between them. With manual operation the switching is so limited that the chances of connecting circuits with echo suppressors in tandem are small and it has been practicable to apply echo suppressors on the basis of round-trip delay of the individual circuits. However, with dial operation it will be possible to establish connections which are long enough to require an echo suppressor but which are composed of circuits each too short to require an echo suppressor based on its round-trip delay. For example, an echo suppressor would not normally be used on a 500-mile carrier circuit, but if eight such circuits were connected in tandem giving a total length of 4,000 miles, an echo suppressor would be imperative, if over-all loss is not to be excessive.

It is not practicable to take care of this problem merely by reducing the delay time at which an echo suppressor is applied, since if this were done it is conceivable that eight circuits each with an echo suppressor might be connected in

Table V. Typical Losses

Type and Length of Trunk	VNL, Decibels	TNL, Decibels
50-mile *N1* carrier	0.5	4.5
50-mile 2-wire *H-88*	1.9	5.9
200-mile 4-wire *H-44*	2.4	6.4
500-mile *K* carrier	1.4	5.4

tandem. It has been necessary, therefore, to establish more or less arbitrary rules to insure at least one echo suppressor on long connections and to make it very improbable that more than two will be encountered. In general, these rules specify that echo suppressors will be placed on:

1. All RC-NC circuits.
2. All RC-RC circuits.
3. On high-usage group circuits when the desired losses cannot be met without them.

Our ideas as to when suppressors of item 3 will be required may change with the trend from voice-frequency towards high velocity carrier circuits. Experience will be a valuable guide, for it is not likely that an intolerable situation will build up overnight and without casting some shadow of coming echo; and the echo suppressor, being a discrete equipment unit, can be installed after needed without appreciable lost motion or additional cost.

Allocation of Facilities

If the intertoll plant were homogeneous the over-all problem would be solved at this point by designing each circuit in accordance with the preceding. But the plant is not homogeneous; it consists of everything from loaded voice-frequency circuits to circuits on microwave radio with VNLF's ranging from 0.03 to 0.0014. It is therefore necessary to allocate these facilities among different circuit groups in such a way that as far as practicable the higher performance facilities are used in the more demanding parts of the network.

As an aid to allocating facilities, charts like Fig. 5 are used. This figure shows ranges of losses within which circuits in different parts of the network are expected to fall. The losses shown there are exclusive of S, which must be added, as indicated before, at both ends of each connection. It should be emphasized that these losses are not limits in the usual sense, neither are they attempts to divide up over-all losses among circuits. They simply help in allocating facilities in the nonhomogeneous plant among different circuit groups. As the use of carrier is extended, the plant will become more homogeneous and the need for such charts will gradually disappear.

Fortunately, from the transmission standpoint, while the machines will set up a wide variety of connections, the routing patterns will be rigidly controlled. Thus, it is practicable to know for each circuit group the maximum number of other circuits with which it can be used in tandem. The lower velocity circuits, 2-wire circuits, narrow-band circuits, etc., can, within practical limits, be allocated to groups which have relatively easy requirements.

Two-Wire Switching and Other Considerations

Present views are that even the ultimate plan will involve 2-wire switching at many points, mainly at the smaller switching points where the amount of traffic will not support the cost of complete automatic alternate routing features. This will cause some additional complication, for each such point introduces another source of echo, that which occurs because the capacitance and resistance of the office cabling reduces the balance obtainable when two intertoll trunks are switched together. The effect of such switching on VNL's can be cared for by adding appropriate loss increments which will be small if a careful job of impedance matching is done and if the distances from the toll terminal equipment to the switches is held within bounds.

No increment is added if the return loss for about 84 per cent of the circuits in the group is 24 decibels or more. These increments increase to about 0.2 decibel for a return loss of 20 decibels, 0.4 decibel for 18 decibels, and so on. They are added to VNL of circuits between a 2-wire switching point and a 4-wire switching point, and of circuits between two 2-wire switching points. Impedance matching is usually accomplished by adding capacitance across the compromise network and in some cases across the shorter cable runs in an office.

All circuit losses referred to in this paper are 1,000-cycle values, that is, no allowance is made for the effects of noise and frequency distortion. Careful design, layout, and co-ordination of individual transmission systems are depended on to keep noise within proper bounds; and all new carrier systems going into the plant have transmitted bands wide enough to require no assignment of distortion transmission impairment (DTI). Circuits having excessive noise and those circuits with large DTI's are earmarked for improvement by any means that may come along. But beyond this, frequency distortion does not enter into VNL calculations since it cannot be offset by reducing circuit losses without encountering trouble from the echo or other standpoint.

While only circuit design is considered in this paper, it is evident that the success of the whole plan also depends on how closely circuit losses are maintained. This is important from two aspects:

1. The expected variations determine the allowance which must be made in the assigned loss. As indicated previously, it is expected that an allowance of 0.4 decibel per link will be adequate for the near future and it is hoped that as time goes on this figure can be reduced.
2. A more important factor is that unless circuit losses are maintained fairly precisely, large positive or negative excess losses can be accumulated on multiswitched connections. Avoidance of such large excesses is particularly important with dial operation since detection and avoidance of unsatisfactory transmission conditions by operators will be much less effective.

The maintenance problem is at least as complex and difficult as the design problem, but it is beyond the scope of this paper.

Summary

For the particular conditions in the Bell System, an equation has been set up to give adequate approximations of the lowest practicable loss for practically all intertoll trunks as follows

$$\mathrm{VNL} = \mathrm{VNLF} \times L + A + B$$

where

VNL = Via net loss (decibels) of the trunk
VNLF = Via net loss factor, that is, a factor which depends on and is appropriate to the type of facilities used in the trunk
L = Length in miles
A = Design allowance for expected variations of circuit loss in service, 0.4 decibel
B = Amount to be added if 2-wire switching is used; the magnitude depends on the passive return loss obtainable on such connections at the 2-wire switching office

At each end of the connection a loss of 2 decibels ($S=2$) is added by appropriate means, as discussed earlier.

Conclusion

Let it be emphasized that this discussion has been largely of planning for the future in all that has preceded, for the switching plan as outlined is growing, and it will be a couple of years before much complex automatic alternate routing is accomplished. It would be surprising if the project should escape growing pains and change of ideas as the plan develops. The author is confident, however, that the plan is sound economically and transmission-wise; and flexible enough to adapt itself to further developments and experience.

References

1. AUTOMATIC SWITCHING FOR NATION-WIDE TELEPHONE SERVICE, A. B. Clark, H. S. Osborne. *AIEE Transactions*, vol. 71, pt. I, Sept. 1952, pp. 245–48.

2. FUNDAMENTAL PLANS FOR TOLL TELEPHONE PLANT, J. J. Pilliod. *Ibid.*, pp. 248–56.

3. NATION-WIDE NUMBERING PLAN, W. H. Nunn. *Ibid.*, pp. 257–60.

4. AUTOMATIC TOLL SWITCHING SYSTEMS, F. F. Shipley. *Ibid.*, pp. 261–69.

◆

Discussion

R. C. Crawshaw (New York Telephone Company, Albany, New York): This discussion is not so much intended to be a direct comment on the material in Mr. Huntley's paper, but more to bring out its application to a going telephone plant, a matter that is of very practical interest in relation to the subject.

It is desirable at the outset to emphasize one point but briefly touched upon in the paper. The development of the national toll-switching plan employs an entirely new concept of switching with new switching instrumentalities at the hub of things. Of necessity, these will quite completely displace the older switching equipment. In this respect the switching end of the business ultimately gets an almost entirely new start—a clean break, as it were—except perhaps in a minor way at the outer edges.

On the other hand, the transmission system is not concentrated in the form of equipment and plant investments within a few building locations, as is the switching equipment. It is obviously impractical to take the replacement approach, since it consists of wire lines, cables (aerial, underground and buried), repeater or amplifier stations, radio system and coaxial cables, etc., scattered over the country covering tremendous route mileage. It is thus harder to reach and to change to conform to the new strains put on it to meet the more exacting transmission requirements Mr. Huntley has described. As a very simple example of a most complex problem, the desired transmission results can be upset by a poorly located loading point buried under the plains of Missouri or in the rugged terrain of New York's Catskill Mountains. This coil may have been adequately located to meet the structural or electrical circuit balances and permit the repeater gains and echo control required under the old switching plan, but it might not be adequate under the new one.

From the transmission viewpoint, therefore, it was of considerable importance to know that the large investments in existing plant built up over many years could with a minimum of expense be fitted in or adapted.

Mr. Huntley has given a clue to one part of the threefold answer to this very practical problem. In this switching plan some circuits will not be required to meet maximum switching demands. As described in the Minneapolis symposium on switching, these are the so-called high usage groups and circuits for certain other uses, as opposed to the final groups. Judicious assigment of facilities and routing of message traffic results in preserving the more flexible ones for final group use. This requires much study and day-to-day attention.

The second part of the answer lies in our advance planning. New instrumentalities, known to be in the development stage, such as *N* carrier systems, were fitted into current route relief plans long before actual availability. In some cases circuits to provide for current growth were temporarily misrouted to stall off relief on a given route until new facilities would be available that would more adequately meet these future transmission requirements than those immediately at hand. This was done on backbone routes destined to carry the important final groups on the switching plan, with the most important ones getting first attention. An example of this was in respect to the important long-haul routes between Buffalo, Fredonia repeater station, and Jamestown in New York State, where by careful planning and facility matching, some new requirements during the last few years were met by using roundabout routes until an *N* carrier development could be made available to meet both current and future demands.

The third part of this answer lies in the application of more exacting maintenance requirements on all toll facilities. Procedures to do this have been adopted and are working out very well. Time does not permit a detailed review of this other very practical side of the transmission problem.

It perhaps suffices to say that the over-all result of applying these new transmission concepts to the telephone plant has been good. What looked like another case of a good theory but hard to practice is turning out to be sound in both theory and practice. Practically all of the existing intercity plant is being put to useful purpose by advance planning and careful study.

Generator Stator Copper Temperature Indicator

A. L. BROWNLEE
ASSOCIATE MEMBER AIEE

H. E. BROWN
ASSOCIATE MEMBER AIEE

A MEANS of measuring the copper temperature of a high-voltage generator conductor is described. A temperature potentiometer measures the temperature drop through the insulation. The temperature drop through the insulation added to the temperature of the outer surface of the insulation is the temperature of the copper.

The resistance temperature detector (RTD) is commonly used as the temperature guide for operation of the stators of large high-voltage generators. These detectors indicate the temperature at their location in the stator, usually between stator conductors or in the bottom of the slot. In either case they are in contact with the outer surface of the conductor insulation and measure the average surface temperature and not the copper conductor temperature. To obtain the copper temperature, the temperature drop through the insulation must be added to the temperature of the RTD.

The temperature drop through the insulation is variable, depending on the generator load, iron temperature, and machine cooling. It can be estimated from data on pilot generators of similar design but preferably should be measured.

A temperature potentiometer is used to measure the drop through the insulation. Fig. 1 shows a temperature potentiometer in which two thermal insulators are in contact. T_c, T_1, and T_2 are temperatures at surfaces indicated; R and r are corresponding thermal resistances. When there is no heat flow through the insulation $T_c=T_1=T_2$. When there is a difference in temperature between T_c and T_2, heat flows through the insulation and the following relationship holds: $T_c=R/r\quad(T_1-T_2)+T_1$. With the ratio of thermal resistances R/r

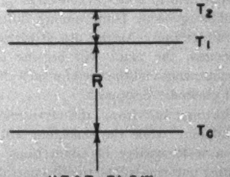


Fig. 1. Temperature potentiometer

Paper 53-294, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 16, 1953; made available for printing April 30, 1953.

A. L. BROWNLEE and H. E. BROWN are with the Commonwealth Edison Company, Chicago, Ill.

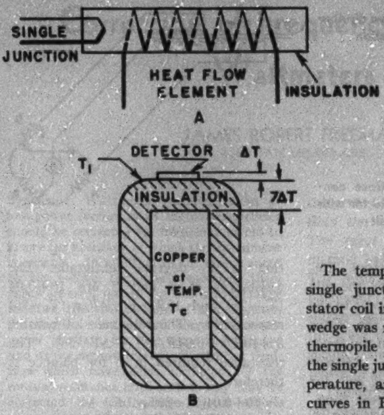


Fig. 2. Thermal electric conductor temperature detector and method of mounting

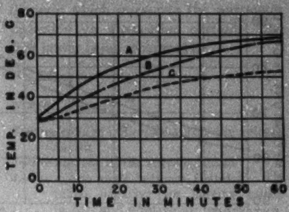
chosen equal to n, a whole number, the equation reduces to $T_c = n\ (T_1 - T_2) + T_1$.

If a thermopile of n couples is formed with one set of junctions at T_1 and the other set at T_2, an electromotive force is produced that is proportional to $n\ (T_1 - T_2)$. A single thermal junction at T_1, in conjunction with a compensated junction at the measuring instrument, produces an electromotive force proportional to T_1. The thermpile connected in series with the single junction and the compensated junction at the instrument produces an electromotive force proportional to $n(T_1 - T_2) + T_1$ or T_c.

This temperature potentiometer is the method used in this device. T_c is the temperature of the copper, which cannot be measured directly but is the quantity to be determined. T_1 is the temperature of the exterior of the insulation and is measured. T_2 is the temperature on the remote side of the thermopile and in combination with T_1 establishes the temperature drop through the insulation.

A thermopile and single junction were mounted on a strip of pressed mica 0.030 inch thick, 2 inches long, and $^1/_2$ inch wide with thermal resistance r, and installed in the slot wedge of a 30,000-kw 9,000-volt 25-cycle generator. The ratio R/r as determined by test was seven. Seven thermocouples were used in the thermopile, with the additional single junction for measurement of the surface temperature. The construction of the thermopile and the single junction is shown in Fig. 2(A) and its position on the stator coil is shown in Fig. 2(B).

Fig. 4 (rigt). Conductor thermal electric temperature detector response as a function of time



The temperature, as indicated by the single junction, on the surface of the stator coil insulation in contact with slot wedge was recorded. The output of the thermopile was added to the output of the single junction to give the copper temperature, and was also recorded. The curves in Fig. 3 are plotted from data recorded with the generator in service carrying load. These curves show the variation in temperature of the surface of the insulation and the copper for various loads on the machine.

Laboratory tests were made to check the time rate of response and the expected accuracy of the detector. An armature coil removed from a 60-cycle 12,000-volt generator was used for these tests. The copper temperature was measured by a thermocouple attached to the copper and the thermopile with a thermal junction for surface temperature was taped in contact with the outer surface of the insulation. In Fig. 4 the temperature of the copper, curve A the copper temperature as indicated by the thermopile, curve B, and the temperature of the surface of the insulation, curve C, are plotted against time. In a 40-degree rise in an hour, the indicated copper temperature approached true temperature within 2 degrees, the indicated temperature lagged behind the true temperature a maximum of 8 degrees. When the 8-degree lag occurred, the indication lagged the true temperature 10 minutes in time. In this test five couples were used in the thermopile. The curves show that the temperature drop through the insulation as indicated by the thermal pile was too small. This means that the ratio of R/r was greater than five. The error could have been reduced by slightly modifying the thermopile.

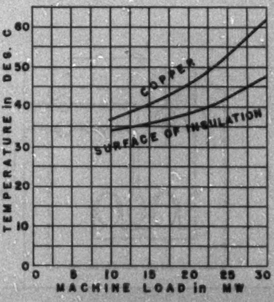


Fig. 3. Copper and surface insulation temperatures of stator winding as a function of machine loading

$T_c = R/r\ (T_1 - T_2) + T_1$ is the equation of a straight line and assumes the temperature distribution to be uniform throughout the insulation. To check the distribution, thermal junctions were embedded in the insulation of an armature coil. The conductor was heated by circulating current. The curves of Fig. 5 show this distribution. Curve A is the temperature measured through the insulation on the top edge of the coil. Curve B is the temperature measured at the same junctions after copper plates had been clamped to the two sides of the coil to increase side surface cooling. The increased cooling lowered the insulation and conductor temperature with little change in the temperature distribution.

The temperature distribution as shown by the curves of Fig. 5 is not uniform. A detector installed on this edge of the coil would indicate a lower temperature than the true temperature of the copper if the ratio R/r is determined from the thermal resistances of the insulation and the thermal pile. However, if the ratio is corrected for the unequal temperature distribution the conductor temperature can be indicated with negligible error.

No information is available on the change in thermal resistance of armature insulation with age. From examination of the slot portion of many coils removed from generators after many years of service, no large change is expected in the ratio of R/r where the insulation in the thermopile is similar to the coil insulation. The change in the ratio R/r only

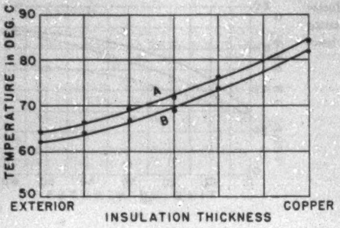


Fig. 5 (left). Temperature distribution in coil insulation

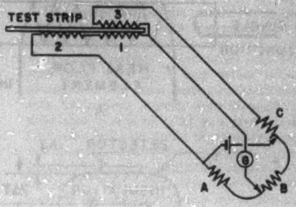


Fig. 6 (right). Resistance conductor temperature detector with measuring circuit

affects the term $R/r\ (T_1-T_2)$ which is of the order of 20 degrees centigrade maximum.

The resistance temperature detector method now generally used can be modified to give copper temperature rather than the exterior of the insulation. The measurement is independent of the lead resistance, provided that the resistances of the leads are equal. In the modification, equal resistances are added in the exploring coil leads. These resistances are mounted on opposite sides of a test strip of thermal resistance r. The measuring circuit and details of the detecting device are shown in Fig. 6. Resistors 1, 2, and 3 are all copper resistors. Resistors 1 and 2 are mounted on the side of the strip that is placed adjacent to the exterior of the insulation of the stator coil. Resistor 3 is mounted on the opposite side of the thermal strip. Resistor 1 is the conventional exploring coil, for instance, 10 ohms at 25 degrees centigrade. Resistors 2 and 3 are equal to $10R/r$ at 25 degrees centigrade, where R/r is the ratio of the thermal resistances of the stator coil insulation to the thermal resistance of the detector strip.

Assume that the temperature difference across the strip is ΔT and that the temperature of the exterior of the insulation is T_1 degrees centigrade. The resistances of 1 and 2 are $10+0.04\ (T_1-25)$ and $10R/r+0.04R/r\ (T_1-25)$ respectively. The resistance of resistor 3 is $10R/r+0.04R/r\ (T_1-\Delta T-25)$. The resistance required in C for balance is therefore

$$
\begin{aligned}
C&=10+0.04(T_1-25)+10R/r+0.04R/r(T_1-25)-[10R/r+0.04R/r(T_1-\Delta T-25)]\\
&=10+0.04(T_1-25)+0.04R/r\Delta T\\
&=10+0.04(T_1+R/r\Delta T-25)
\end{aligned}
$$

The bridge will therefore indicate a temperature R/r times the differential across the strip higher than T_1. This is the temperature of the copper.

◆

Discussion

Herman Halperin (Commonwealth Edison Company, Chicago, Ill.): As indicated in another discussion,[1] great attention should be given to the temperature of the conductors themselves in determining allowable loads. Information is needed on average conductor temperatures in connection with the movement of the coils in the slots, and is needed on hottest-spot conductor temperatures in connection with the allowable total temperature for the insulation.

With the advent of hydrogen cooling, the matter of determining the conductor temperatures accurately has assumed increased importance. This is due to the fact that with hydrogen-cooled generators, especially at elevated pressures, a large portion of the temperature rise between the ambient and conductor occurs across the insulation of the armature coil. For air-cooled machines, the thermal drop across the armature coil insulation is relatively small compared to the total temperature rise.

The manufacturers have been aware of this situation for some time. It is gratifying to learn that they are conducting investigations to obtain more information on the actual total temperatures of the conductors in generators.

The authors have developed some ingenious schemes for obtaining more accurate ideas of conductor temperatures of generators after they go into service. They indicate, however, that the applications of their methods require care in order to avoid erroneous conclusions. It would be of interest to see what others find in trial use of the authors' scheme. Also, it would be of interest to see how results obtained by the authors' scheme methods compare to the findings of the manufacturers based on the methods used in the manufacturers' present investigations.

REFERENCE

1. Discussion by Herman Halperin of LOADING OF HYDROGEN-COOLED GENERATORS AT ELEVATED GAS PRESSURES, D. S. Snell. *AIEE Transactions*, vol. 69, pt. I, 1950, p. 199.

Correction of Frequency Errors in Wattmeters

JAMES ROBERT FREEMAN
ASSOCIATE MEMBER AIEE

Synopsis: Wattmeter readings taken at low power factors or at higher frequencies should be corrected for frequency errors to insure the accuracy of which the instrument is capable. In addition to the error caused by the voltage circuit self-reactance, there exists an error due to the mutual reactance between the two circuits of a wattmeter. A practical method of correcting wattmeter readings for this error is presented and its use demonstrated. Wattmeter readings which have been corrected for the mutual reactance error in this way may be corrected for the voltage circuit self-reactance error by using the customary tables or correction factors in the conventional manner. Curves are shown of the two frequency errors as functions of the frequency and the deflection for a typical commercial low power-factor wattmeter.

ELECTRODYNAMOMETER wattmeters are generally accepted as the most satisfactory type of indicating instruments for the measurement of power at low frequencies. These instruments may have accuracies to within 0.1 per cent of full-scale deflection when used within reasonable limits of frequency and power factor. Most portable types of commercial precision electrodyanamometer wattmeters have a guaranteed limit of error of 0.25 per cent of full-scale deflection on frequencies up to 133 cycles per second, but because of inherent errors this accuracy is not maintained at higher frequencies or with very low power factors. The inherent errors referred to may be attributed to two things: the self-reactance of the voltage circuit, including distributed capacitance; and the mutual reactance between the voltage coil and the current windings. The existence of eddy currents within an instrument will also cause an error in indication. Experiments have revealed this error to vary at a rate slightly higher than the first power of the frequency. The error may be determined by a method based on bridge measurements.[1] The results of such measurements on a particular commercial wattmeter have shown that the error is 0.4 per cent of full-scale deflection at 1,200 cycles per second, and is proportional to the 1.3 power of the frequency.

The error caused by the voltage circuit self-reactance is well understood, and correction tables are available which give factors by which wattmeter readings may be multiplied to correct for this error.[2,3] For measurements at frequencies below about 600 cycles per second, these corrections are usually sufficient to insure the accuracy of which a wattmeter is capable. At higher frequencies, however, the error caused by the mutual reactance may be significant also, and it then becomes necessary to have a correction for it. At these higher frequencies attention should be given to this problem.

Effect of Frequency

Not considering the effect of eddy currents, the indication of a wattmeter at a particular frequency depends only upon the product of the currents in the two wattmeter circuits and the cosine of the phase angle between them. Fig. 1 is a schematic diagram of an electrodynamometer wattmeter and Fig. 2 is a phasor diagram which interprets the wattmeter operation on alternating current. The true power being measured is

$$P=EI\cos\theta \tag{1}$$

where

E = the voltage impressed across the voltage circuit
I = the current in the current circuit
θ = the phase angle between E and I. θ is taken negative for lagging power factors and positive for leading

The effect of the mutual reactance between the two wattmeter circuits is to induce a voltage into the voltage circuit. This voltage is represented by the phasor $I\omega M$ and is in quadrature with the current I. Therefore the impressed voltage is given by the phasor equation

$$E\underline{/\theta}=I'Z\underline{/\bar{\theta}}+I\omega M\underline{/-90^\circ} \tag{2}$$

where

I' = the current in the voltage circuit
Z = the self-impedance of the voltage circuit

With direct current

$$E=I'R \tag{3}$$

and if a wattmeter reads correctly on direct current, it is apparent that the instrument is calibrated so as to make the voltage measured by the instrument equal to $I'R$.

With alternating current, the current flowing through the voltage circuit will lag the impressed voltage by the angle α which is the phase angle of the voltage circuit impedance. Therefore from Fig. 2 the wattmeter indication with alternating

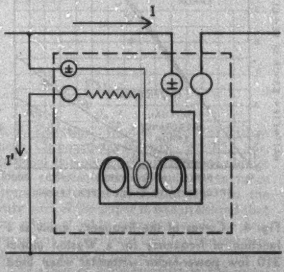


Fig. 1. Schematic diagram showing the two electric circuits of an electrodynamometer wattmeter

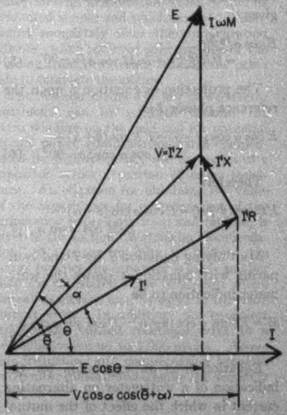


Fig. 2. Voltage phasor diagram which indicates the operation of a wattmeter on alternating current

---

Paper 53-295, recommended by the AIEE Instruments and Measurements Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted June 12, 1952; made available for printing April 29, 1953.

JAMES ROBERT FREEMAN, formerly with the University of California, is with the Massachusetts Institute of Technology, Cambridge, Mass.

The author wishes to acknowledge the assistance and instruction received from Wilson S. Pritchett during the preparation of this paper. Acknowledgement is also made to Prof. Charles F. Dalziel, under whose supervision the early investigations of this problem were carried out.

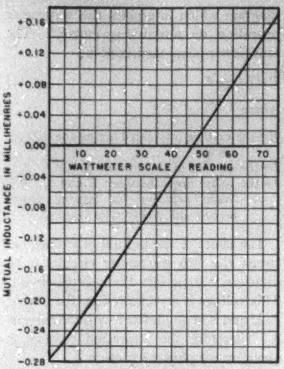


Fig. 3. Curve showing the value of the mutual inductance between the two circuits of a Weston model 310 low power-factor wattmeter as a function of the wattmeter deflection. Current windings connected in multiple (5-ampere connection). For the series connection (2.5 amperes) the values of mutual inductance are to be multiplied by 2. The wattmeter voltage circuit is connected uncompensated for voltage circuit loss to minimize the mutual inductance

current is

$$W=I\,I'R\cos(\bar{\theta}+\alpha) \tag{4}$$

The projections $E\cos\theta$ and $I'R\cos(\bar{\theta}+\alpha)$ on the reference phasor I show the relative difference between the true power component of voltage and the one observed by the wattmeter. Multiplying equation 2 by the phasor quantity $\cos\alpha\,\underline{/\alpha}$ gives

$$E\cos\alpha\,\underline{/\theta+\alpha} = I'R\,\underline{/\bar{\theta}+\alpha}+I\omega M\cos\alpha\,\underline{/\alpha-90^\circ} \tag{5}$$

The projection of equation 5 upon the reference phasor I is

$$E\cos\alpha\cos(\theta+\alpha)=I'R\cos(\bar{\theta}+\alpha)+ I\omega M\cos\alpha\cos(\alpha-90^\circ) \tag{6}$$

Therefore

$$I'R\cos(\bar{\theta}+\alpha)=E\cos\alpha\cos(\theta+\alpha)- I\omega M\cos\alpha\sin\alpha \tag{7}$$

Multiplying equation 7 by I and comparing with equation 4 shows the wattmeter indication to be

$$W=EI\cos\alpha\cos(\theta+\alpha)-I^2\omega M\sin\alpha\cos\alpha \tag{8}$$

Equation 8 is an expression for the indication of a wattmeter on alternating current in which the effect of the mutual reactance has been isolated as an independent term. From equation 8 and Fig. 2 it can be seen that without mutual reactance the wattmeter indication would be

$$W'=EI\cos\alpha\cos(\theta+\alpha) \tag{9}$$

Correction for Voltage Circuit Self-Reactance

If the effect of the mutual reactance is insignificant, or if the value of the wattmeter reading W has been adjusted to correct for the effect of the mutual reactance, comparison of equations 1 and 9 reveals that by multiplying W' in equation 9 by

$$\cos\theta/\cos\alpha\cos(\theta+\alpha) \tag{10}$$

the value of the true power, as given in equation 1, is obtained. The values of $\cos\theta/\cos(\theta+\alpha)$ can be found by entering tables with the values of α and $\cos(\theta+\alpha)$.[2,3]

Correction for Mutual Reactance Between Current Windings and Voltage Coil

For measurements where the effect of the mutual reactance should not be ignored, it is possible to correct wattmeter readings for this error by adding a correction term ΔW, to the wattmeter reading W, which gives an adjusted reading

$$W'=W+\Delta W \tag{11}$$

If W' is the wattmeter reading corrected for the effect of the mutual reactance (equation 9), then from equation 8 it follows that

$$\Delta W=I^2\omega M\sin\alpha\cos\alpha \tag{12}$$

As shown in the Appendix, the expression for ΔW may be simplified to the form

$$\Delta W=0.11fMI^2\alpha \tag{13}$$

where

f = the frequency in cycles per second
M = the mutual inductance between the current windings and the voltage coil expressed in henrys. M may be either positive or negative
I = the current being measured by the current circuit expressed in amperes
α = the phase angle of the voltage circuit self-impedance expressed in decimal parts of a degree

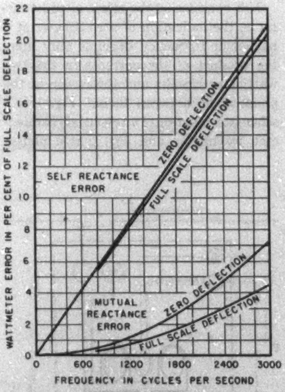


Fig. 4. Curves of the frequency errors as a function of frequency for a Weston model 310 low power-factor wattmeter when used on the 75-volt range with rated voltage and current. Rating: 75 volts, 2.5/5 amperes, 37.5/75 watts

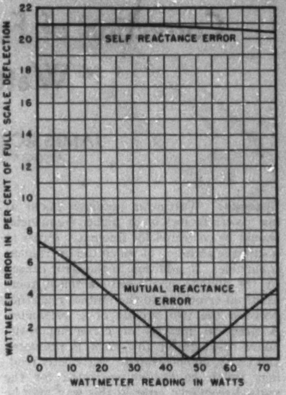


Fig. 5. Curves of the frequency errors at 3,000 cycles per second as a function of wattmeter reading for a Weston model 310 low power-factor wattmeter when used on the 75-volt range with rated voltage and current. Rating: 75 volts, 2.5/5 amperes, 37.5/75 watts. For 2.5-ampere connection wattmeter reading is divided by 2

Applying the Correction

In applying the correction for the mutual reactance, it is obvious that one must first know the value of the mutual inductance M. The value of the mutual inductance depends upon the orientation of the voltage coil with respect to the current windings. This means that the value varies with the indication of the wattmeter. When an instrument is connected for up-scale deflection, M is negative at the lower end of the scale and positive at the upper end, having a value of zero somewhere near mid-scale.

The mutual reactance between the two

circuits of a wattmeter may be determined by measuring the voltage induced in one circuit by a known current flowing through the other circuit. If this is done for several positions of the pointer, a curve such as the one in Fig. 3 can be drawn. The voltage should be measured with a high impedance device such as a vacuum-tube voltmeter or an oscilloscope. Fig. 3 is a curve of the mutual inductance versus wattmeter deflection for a Weston model *310* low power-factor wattmeter.

As an example of the use of equation 13 and the curve in Fig. 3, consider this wattmeter to be used at a frequency of 1,200 cycles per second. Assume the wattmeter reading to be $W=22.5$ watts, and the current measured to be $I=4.97$ amperes. For this wattmeter, $\alpha=1.13$ degrees at 1,200 cycles per second. From the curve of Fig. 3, $M=-0.15$ millihenry. Setting these values into equation 13 gives $\Delta W=-0.6$ watt. Therefor the wattmeter reading corrected for the effect of mutual reactance is $W'=21.9$ watts. If the correction for the mutual reactance were not applied, an error in the reading of 2.7 per cent would exist due to this effect. This error of 0.6 watt would be three times that of the guaranteed limit of error for the instrument.

Comparison of Frequency Errors

Figs. 4 and 5 are curves of the frequency errors for a Weston model *310* low power-factor wattmeter when used on the 75-volt range with rated voltage and current. Fig. 4 shows how the two errors vary with frequency. The mutual reactance error varies as the square of the frequency. This is so because ΔW is proportional to the frequency times α, α itself being proportional to the frequency since it is a small impedance angle. The fact that the mutual reactance error is proportional to the square of the frequency, whereas the error caused by the voltage circuit self-reactance is proportional only to the first power of the frequency, emphasizes the importance of correcting for the mutual reactance error when measurements are made at higher frequencies.

The curves of Fig. 5 show how the two frequency errors vary with the wattmeter deflection with rated voltage and current applied at 3,000 cycles per second.

Conclusions

1. The error in indication caused by the mutual reactance between the two circuits of a wattmeter is proportional to the square of the frequency and is given by equation 13.

2. The mutual reactance error may be corrected for by use of the correction term ΔW as given in equation 13. Wattmeter readings corrected for the mutual reactance error in this way may then be corrected for the voltage circuit self-reactance error by the usual methods.

Appendix. Simplification of Correction Term for Mutual Reactance Error

Equation 12 is an exact expression for the error due to the mutual reactance between the two circuits of a wattmeter

$$\Delta W=I^2\omega M\sin\alpha\cos\alpha \tag{12}$$

Since

$$\omega=2\pi f \tag{14}$$

and

$$2\sin\alpha\cos\alpha=\sin 2\alpha \tag{15}$$

equation 12 may be written

$$\Delta W=\pi fMI^2\sin 2\alpha \tag{16}$$

Furthermore, since α is a small angle

$$\sin 2\alpha=(\pi/180°)2\alpha \tag{17}$$

where α is expressed in decimal parts of a degree. Substituting equation 17 into equation 16 gives

$$\Delta W=(\pi^2/90)fMI^2\alpha \tag{18}$$

Therefore

$$\Delta W=0.11fMI^2\alpha \tag{13}$$

References

1. Correction of Frequency Errors in Electrodynamometer Wattmeters, J R. Freeman. *Thesis*, University of California Library, Berkeley, Calif., 1951.

2. Electric Power Metering (book), A. E. Knowlton. McGraw-Hill Book Company, Inc., New York, N. Y., 1934, pp. 1–15, 36–39.

3. Manual of Instrument Transformers. General Electric Company, Schenectady, N. Y., 1950, pp. 46–47.

4. Electrical Measurements (book), F. K. Harris. John Wiley and Sons, Inc., New York, N. Y., 1952.

◆

Discussion

R. F. Estoppey (Weston Electrical Instrument Corporation, Newark, N. J.): Mr. Freeman has presented an accurate picture of the self-reactance and mutual-reactance errors that occur in wattmeters. In the early part of the paper it was indicated that the effect of eddy currents would be neglected but it is believed that further mention should be made of this error.

The effect of eddy currents in the metal parts surrounding the moving element causes the field coil flux to lag the field coil current by a small angle. Tests on the Weston model 310 low power-factor wattmeter cited in the paper indicate that this angle is approximately 0.4 degree at 1,200 cycles. At rated voltage and current which corresponds to 20 per cent power factor, this angle will cause an error of 3.5 per cent of full-scale deflection.

From the curves of Fig. 4 it can be seen that the mutual-reactance error is approximately 1 per cent at 1,200 cycles and the self-reactance error is approximately 8 per cent at 1,200 cycles. Since the error due to the eddy currents is appreciable, it is felt that this error should be considered if the curves of Fig. 4 are used to correct wattmeter indications.

Since the self-reactance error causes the moving coil current to lag the applied voltage and the eddy current error causes the field coil flux to lag the field coil current, the effect of these angles is to cancel each other and if they were equal there would be no error from these two sources. The combined error, due to these two causes, then becomes the difference between 8 per cent and 3.5 per cent, or 4.5 per cent.

The Weston model *310*, form *2* low power-factor wattmeter, chosen for this study, was designed for 60-cycle use. At this frequency, and up to 125 cycles, the compensation for the power taken by the potential system simplifies the taking of accurate readings under a variety of conditions. This compensation consists of an additional winding on the field coils, turn for turn with main winding, and connected in series with the moving coil; it can be connected in at will by means of a switch so that the user of the instrument may use the compensation feature or omit it.

The compensating winding is closely coupled with the field coils, both inductively and capacitively, and adds additional mutual-inductance error when the instrument is used with potential circuit compensation.

Due to the capacitive coupling between windings, part of the main moving coil current can flow from the compensating winding directly into the field coil winding. This causes an error in indication which cannot be evaluated simply and which cannot be eliminated completely since the compensation selector switch removes only one side of the winding from the circuit, leaving the other side to complete the circuit.

The Weston model *310* low power-factor wattmeter can be supplied without the extra winding on the field coil for potential circuit power compensation, and compensated for high-frequency use by means of a capacitor across a portion of series resistance. An equation for obtaining the value of the capacitance for compensation of the inductance errors both fixed and mutual, is given in the section "Wattmeter Compensation" of a paper by John H. Miller.[1] In actual practice the capacitor is made smaller than the calculated value to a point where the self-reactance error is equal to the eddy current error. In addition, the capacitor reduces the mutual reactance error to a negligible value up to 1,000 cycles. Above this frequency, any compensation becomes much more difficult and usually the accuracy is impaired.

Reference

1. Frequency Compensation of A-C Instruments, John H. Miller. *AIEE Transactions*, vol. 70, pt. I, 1951, p. 220.

Wilson Pritchett (University of California, Berkeley, Calif.): In low power-factor measurements the self-reactance correction becomes a term which may also be added to the wattmeter reading. It is convenient to use instead of the phase angle α the ratio of the reactance to the resistance Q of the voltage circuit. Both corrections in terms of Q are $2\pi fMQI^2$ and EIQ. The first due to the mutual reactance is valid for all power factors, and the second is accurate to within 2 per cent for power factors less than 0.2. Both depend upon $Q^2/2$ being negligible compared to unity. This is implied by equation 17 in the paper. This approximation is true for conventional wattmeters up to more than 10 kc where other factors tend to prevent use of such wattmeters for power measurement. The sign of Q is positive for leading power-factor loads, and the sign and magnitude of M depend upon the data of Fig. 3.

A practical method not involving breaking the seal for obtaining these data employs direct current to produce the deflection and alternating current to measure the mutual inductance. Appropriate a-c and d-c sources are connected in series with the current coils, a rheostat, and two ammeters. One of these is a d-c type and the other is a true rms type such as the electrodynamic ammeter. The square root of the difference between the squares of the two readings gives the a-c component to be used in computing M. The vacuum-tube voltmeter is connected to the voltage circuit terminals which are energized by a d-c source in series with a high a-c impedance element. Use of a conventional 32-henry choke coil rated at 40 milliamperes with 60-cycle current has given results hardly distinguishable from those obtained by deflecting the pointer manually.

The a-c and d-c supplies for the current circuit should mutually afford a low impedance path for the current of the other. This direct current, if not obtained from a battery, should have a ripple component low compared to the alternating current. The d-c supply for the voltage circuit must either be a battery or must be free from ripple or noise. Grounding the watt-meter terminals marked with the ± symbols helps to eliminate the effect of stray electric fields.

The voltage readings should be corrected for scale error of the vacuum-tube voltmeter and for the loading effect of the choke coil and battery. The latter correction is determined on a percentage basis by making the measurement at zero deflection both with and without the choke and the power supply connected. The resulting mutual inductance curve for uniformly spaced scale divisions is a straight line.

Use of the compensating winding adds a fixed amount of negative mutual inductance to the data. This fixed mutual inductance is four times as much for the series connection as it is for the parallel connection. In low-current wattmeters the fixed mutual inductance of the compensating winding is not only large but it is a much larger proportion of the total mutual inductance than it is in high-current wattmeters. For example, in a 30-watt 1.0/2.0-ampere 75/150-volt instrument with series-connected coils, the fixed mutual inductance is −4.5 millihenrys compared to a maximum of −1.3 millihenrys for the variable mutual inductance. Corresponding values for a 600-watt 20/40-ampere 75/150-volt instrument with parallel coils are −0.003 millihenry and −0.033 millihenry. In the first instrument the fixed component is over three times as large as the maximum value of the variable component. In the second it is less than one-tenth as large.

Fig. 1 shows the connection for power flowing toward the right. After the frequency corrections are applied the true power obtained includes that consumed by the current coils. Changing the connection of the upper voltage terminal from the left to the right of the current circuit causes the power consumed by the voltage circuit to be included instead. Since readings of current in the current coils and voltage across the voltage circuit (both corrected for frequency) are needed (see equations 13, 9, and 1), the power consumed by either the voltmeter or the ammeter is included also. Computation and then subtraction of the appropriate quantities is required as a final step in determining the true power consumed by the load.

James Robert Freeman: The method proposed by Prof. Pritchett of measuring the mutual reactance at several pointer positions without breaking the manufacturer's seal is an excellent one. The ability to do this is of great value to the user of a commercial instrument since the breaking of an instrument seal should be avoided in order to insure confidence in the instrument and not jeopardize a manufacturer's warranty.

Concerning the use of the Q of the voltage circuit instead of the voltage circuit impedance angle α, both equation 13 of the paper and $2\pi fMQI^2$ are equally satisfactory for the mutual reactance error correction. If the prepared tables for the self-reactance error referred to in the paper are not available, it would seem that the use of the term containing Q would be more convenient. Prof. Pritchett states that the sign of Q is positive for leading power-factor loads. The sign of Q is independent of the power factor since it is the tangent of the impedance angle of the voltage circuit. The exact correction for the self-reactance error is $EIQ \sin\theta$ and the sign associated with the correction is obtained from the sign of the angle θ. Therefore the self-reactance correction term is positive for leading power factors and negative for lagging. The mutual reactance correction term is not dependent on the power factor.

Dielectric Breakdown of Sulfur Hexafluoride in Nonuniform Fields

C. N. WORKS
MEMBER AIEE

T. W. DAKIN
ASSOCIATE MEMBER AIEE

THIS PAPER PRESENTS some of the anomalies in the dielectric breakdown of sulfur hexafluoride SF_6 as functions of pressure and electrode separation in a nonuniform field. The nonuniform field, although difficult to analyze, is of considerable importance from the engineering viewpoint as it is the condition most generally encountered in electrical apparatus. At atmospheric pressure, sulfur hexafluoride has unexpectedly high 60-cycle breakdown voltage for large gaps. In this case, the presence of corona precedes sparkover. This was termed "corona-stabilized breakdown" in a recent paper by Camilli, Gordon, and Plump.[1] This phenomenon will be explained by showing the effect of electrode spacing on the position of the maximum which occurs in the curve of sparkover voltage versus gas pressure. The d-c 60-cycle and impulse-breakdown voltage, as well as corona-onset voltage, will be discussed. The time required for corona stabilization to occur will be indicated.

Effect of Gap Length on 60-Cycle Breakdown

Fig. 1 shows the 60-cycle sparkover voltage of sulfur hexafluoride at 1 atmosphere pressure as a function of electrode spacing. The capacitor plates form an electrode configuration which approaches uniform field conditions. Therefore, this curve serves as a reference. These capacitor plates were 7 inches square, with corners rounded and 1/2-inch-diameter rods welded to the edge. The other electrode configurations produce very nonuniform fields. All these curves cross the capacitor plate curve for spacings of $2^1/_4$ inches or less. It is unusual that as the electrode spacing is increased,

---

Paper 53-235, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City N. J., June 15–19, 1953. Manuscript submitted March 18, 1953; made available for printing April 27, 1953.

C. N. WORKS and T. W. DAKIN are with the Westinghouse Electric Corporation, East Pittsburgh, Pa.

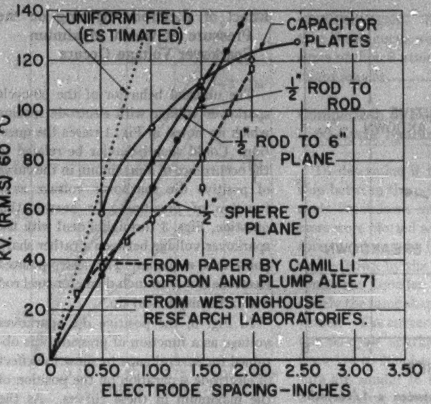


Fig. 1. Sparkover voltage of sulfur hexafluoride as a function of electrode spacing for a number of configurations, at one atmosphere pressure and 25 degrees centigrade

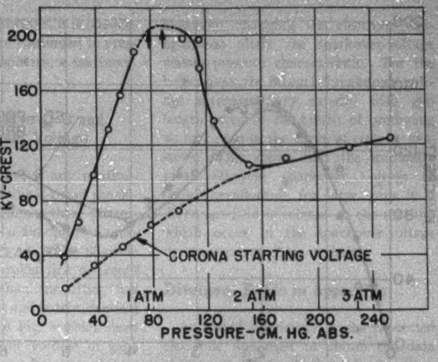


Fig. 2. Positive d-c sparkover voltage of sulfur hexafluoride for a 2-inch gap between 1/2-inch-square rods in 12-inch-diameter steel tank, as a function of pressure

instead of the sparkover voltage levelling off, it may increase at a more rapid rate with spacing, causing an inflection in the curve. Sparkover is preceded by the presence of strong corona which led to the term "corona-stabilized breakdown."

Effect of Pressure on Positive D-C Breakdown

In order to obtain better understanding of the a-c breakdown, the d-c dielectric breakdown was investigated. Fig. 2 shows the positive d-c breakdown voltage of sulfur hexafluoride as a function of pressure for a 2-inch gap between square rods in a steel tank. The grounded tank wall sets up an unsymmetric field. The sparkover voltage passes through a pronounced maximum, and then, at sufficiently high pressure, increases again. This maximum occurs in a region where corona precedes sparkover. The curve of corona-onset voltage increases linearly with pressure. At the higher pressures, sparkover occurs coincident with corona.

Breakdown Mechanisms

The conditions necessary to produce a maximum in the sparkover voltage are a divergent field and a gas with electron-attaching molecules. Such maxima do not occur in gases such as pure nitrogen or helium.[2] Curves showing a maximum in the positive d-c sparkover voltage of sulfur hexafluoride may be found in the literature. Pollock and Cooper[3] investigated a 3-millimeter gap between a point and plane. Results for a 1-centimeter gap were reported by Foord.[2] In the references cited, there is considerable discussion of the mechanisms which may produce such a maximum in the sparkover curve. However, as a background for discussing the experimental results, the following breakdown processes are postulated:

1. In a divergent field, such as is formed by a point-to-plane electrode system, the positive point attracts electrons from weaker parts of the field which, in gases, produces streamers which form long stems with few branches, because electrons are attracted to the advancing streamer tip.

2. With the point negative, the streamers therefrom repel electrons, forming numerous weak short branches, scattered in all directions.

Breakdown from a positive point, therefore, occurs at lower voltage than from a negative point. However, some tests on small gaps[3] have shown an exception to this.

For the point positive, stable corona discharge can occur at low pressures, because the field, due to the space charge at the tip of a positive streamer, is not sufficient to propagate the streamer across the gap. In gases containing electronega-

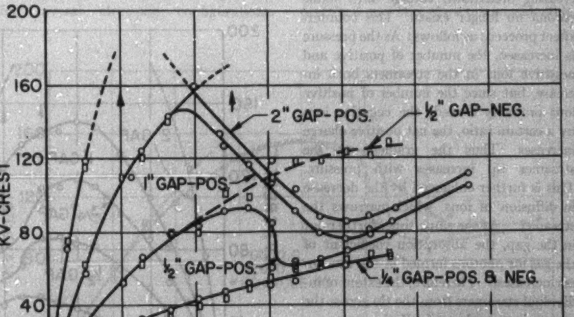


Fig. 3. D-c sparkover voltage of sulfur hexafluoride for several gaps between 1/16-inch-diameter hemispherical point and a 6-inch-diameter plane in 12-inch-diameter steel tank, as a function of pressure

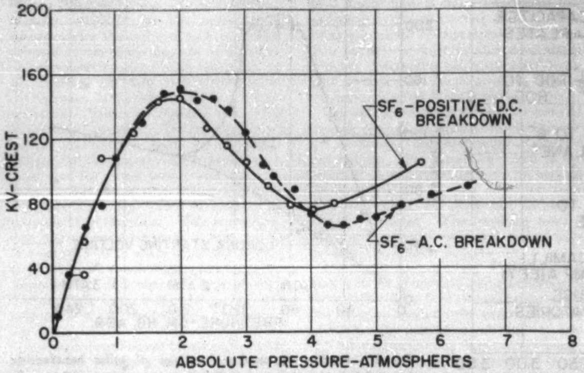


Fig. 4. Sparkover voltage of sulfur hexafluoride for 1-inch gap between a 1/16-inch-diameter hemispherical point and a 6-inch-diameter plane in 12-inch-diameter steel tank, as a function of pressure

tive atoms, such as sulfur hexafluoride, this space charge is reduced in the streamer by the capture of electrons to form negative ions. Photons formed by radiation from the streamer initiate other streamers which diffuse around the point, forming a continuous corona discharge which effectively rounds off the point and makes the applied field less divergent.

In the low-pressure region, increasing the pressure increases breakdown voltage as expected. Coincident with this, a countereffect occurs, which leads to a breakdown maximum, followed by decreasing breakdown voltage until stable corona no longer exists. This countereffect proceeds as follows: As the pressure is increased, the number of positive and negative ions in the streamers both increase, but since the number of positive ions probably exceeds the negative ions by a certain ratio, the net positive charge increases. Thus the gradient at the streamer tip increases with pressure. This is further enhanced by the decrease in diffusion of ions, which narrows the streamer. At the same time, further out in the gap, the absorption coefficient of the gas for photons formed in the streamers increases. Therefore, the extent of individual streamers from the tip across the gap is increased until all the streamers initiating at the point completely cross the gap, producing breakdown coincident with corona. The breakdown mechanism is further complicated by the fact that these positive streamers do not in all cases cross the gap by the shortest possible path. Curved sparks[2,4] have been observed to occur in electronegative gases. These sparks appear to bend around the region occupied by corona. As the pressure is increased, this region is reduced, due to less diffusion. Thus, the path which the curved sparks must travel to cross the gap is shortened. The breakdown depends on the field configuration, which includes the effects of space charges as well as the electrodes. Also, it should be pointed out that a weaker field, acting through a greater distance, can accelerate particles to sufficient energy to initiate breakdown over a circuitous path, as indicated by the curved sparks.

Effect of Electrode Spacing on the Pressure at which Maximum Sparkover Voltage Occurs

The unusual behavior of the 60-cycle sparkover voltage with electrode spacing which we noted in Fig. 1 raises the question: Could such behavior be related to the occurrence of a maximum in the curve of positive d-c sparkover voltage as a function of pressure? To answer this question, Figs. 3 through 6 deal with the sparkover voltage between a rather sharp point and a plane. The sharp-pointed electrode is a 1/16-inch-diameter steel rod with a hemispherical tip.

In Fig. 3, the positive d-c sparkover voltage as a function of pressure was obtained for several gaps to show the effect of electrode separation on the position of the maximum in these curves. As the electrode spacing is increased, this maximum shifts toward lower pressure, and the maximum for larger gaps is broader, encompassing the curves for the small gaps.

Although each succeeding gap is double the former, at higher pressures there is very little increase in the positive sparkover voltage with increase in electrode spacing. Above a certain pressure, the negative strength, indicated by the squares, is much higher than the positive. For example, the negative strength of a 1/2-inch gap at 4 atmospheres pressure is nearly double the positive.

Pollock and Cooper[3] found that, after the maximum was reached, the breakdown voltage dropped off very abruptly, and stated that this was in agreement with

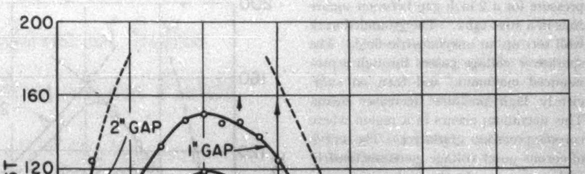


Fig. 5. A-c sparkover voltage for several gaps between 1/16-inch-diameter hemispherical point and 6-inch-diameter plane in 12-inch-diameter steel tank, as a function of pressure

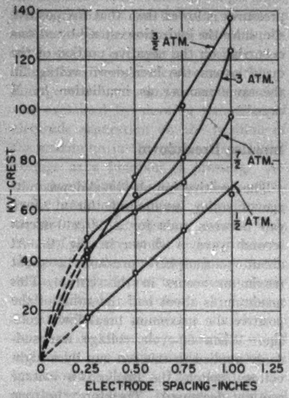


Fig. 6. A-c sparkover voltage of sulfur hexafluoride as a function of electrode spacing between 1/16-inch-diameter hemispherical point and 6-inch-diameter plane, at several pressures

theory. It was found that this is not generally the case. For the 1/2-inch gap in Fig. 3, the positive d-c sparkover voltage decreases very abruptly, whereas, for the 1-inch gap, and especially the 2-inch gap, the decrease is gradual. Also Foord[2] shows the sparkover voltage decreasing slowly over a considerable pressure range. Observation of the curved sparks, mentioned previously, showed that the longer gaps are more productive of curved sparks. Thus, experiment indicates that when breakdown proceeds across the shortest path between electrodes, an abrupt decrease in sparkover voltage may be expected, but when breakdown takes place over a devious route, as indicated by curved sparks, a gradual decrease in sparkover voltage is possible. The performance of a gap is determined more by the divergence of the field than the gap spacing.

As an example, if the curve for a 1-centimeter gap between a very sharp point and a plane, as obtained by Foord,[2] is compared to the curves of Fig. 3, its shape is more nearly like that of the curve for the 1- or 2-inch gap. The breakdown values obtained by Foord are of the same order of magnitude as those for the 1/2-inch gap, Fig. 3. Also, Pollock and Cooper[3] obtained a curve for a 3-millimeter gap between a point of 1/4-millimeter radius and a plane, similar in shape to that for the 1/2-inch gap in Fig. 3. The values of breakdown voltage approach those of the 1/4-inch gap for Fig. 3. The 1/4-inch gap in this figure appears to be approaching uniform field conditions again. However, it is possible that if this curve were continued to pressures of 7 or 8 atmospheres, a maximum might appear.

Comparison of Positive D-C and 60-Cycle Sparkover Voltage

To determine if there is any relationship between the positive d-c and the 60-cycle breakdown voltages, these parameters were plotted as a function of pressure, as shown in Fig. 4. The 60-cycle curve shown by the broken line, although somewhat broader than the other, has closely the same shape and nearly identical values. The curves of Fig. 5 which show the 60-cycle sparkover voltage of gaps from 1/4 to 2 inches as a function of pressure are similar to those for the positive d-c sparkover voltage, except the maxima shown here are somewhat broader than in the former case.

If a pressure in the region of the maxima in the curves of Fig. 5 is selected; for example, 3 atmospheres, and the sparkover voltage as a function of electrode spacing is plotted, the curve labeled 3 atmospheres in Fig. 6 is obtained. This curve has the same shape as the curve for the electrode configuration for 1/2-inch diameter sphere-to-plane shown in Fig. 1. The curve for a pressure of 7/2 atmospheres, Fig. 6, also has an inflection in it. The other curves in this figure are nearly linear. They are similar in appearance to the curves in Fig. 1 for rod-to-rod and rod-to-plane electrode configurations. The curves in Fig. 6 are for several pressures, whereas the curves of Fig. 1 are for several electrode configurations. However, changing the electrode configuration alters the sparkover-voltage versus pressure characteristic. For the larger gaps, the 60-cycle breakdown voltage increases very rapidly with gap length, even to the extent of producing an inflection in the curve at certain pressures. This shows that the unusually great 60-cycle sparkover voltage of sulfur hexafluoride for large gaps in a divergent field is related to the maxima which occur in the sparkover-voltage versus pressure characteristic.

Divergent Fields in Apparatus

As stated previously, a sharp-pointed electrode was used to obtain the data shown in Figs. 3 through 6. Even sharper points were used in the literature cited. Such electrodes were chosen to give fields that could be analyzed. The question may be raised that such electrodes are not representative of divergent fields which may occur in practice. In answer to this possible objection, Fig. 7 is presented, which shows the breakdown voltage between a 1/2-inch square rod and a plane as a function of pressure. Such an electrode arrangement has been used in engineering tests. It should be pointed out that these curves have not only the same shape as the ones referred to previously, but also nearly the same values of sparkover voltage. In the case of Fig. 1, which introduced us to the phenomenon of corona-stabilized breakdown, all electrodes were those which are often used in engineering tests. Therefore, the phenomena which have been described may take place to a considerable

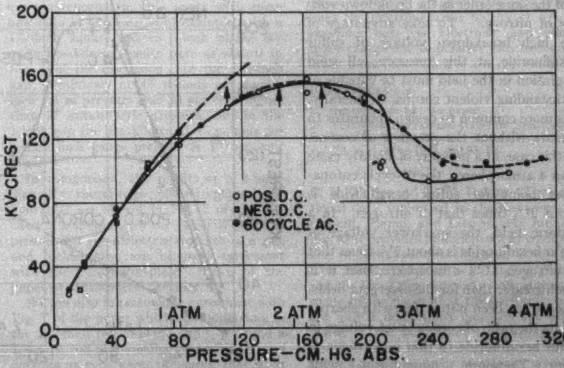


Fig. 7. Sparkover voltage of sulfur hexafluoride for a 1-inch gap between 1/2-inch-square rod and 6-inch-diameter plane in a 12-inch-diameter steel tank, as a function of pressure

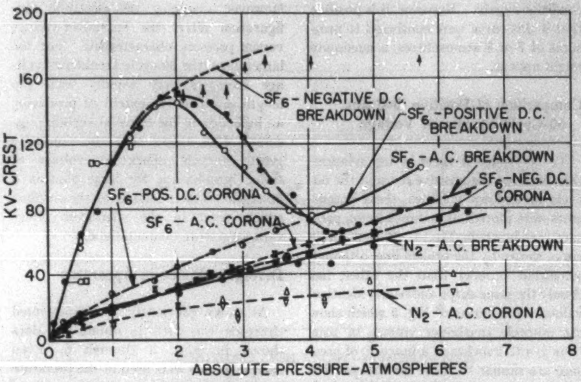


Fig. 8. Sparkover and corona onset voltages of sulfur hexafluoride and nitrogen for a 1-inch gap between 1/16-inch-diameter hemispherical point and 6-inch-diameter plane in a 12-inch-diameter steel tank, as a function of pressure

degree wherever a divergent field exists in electric apparatus.

Relative Strength and Corona Onset in Sulfur Hexafluoride and Nitrogen

In Fig. 8, the sparkover voltage of sulfur hexafluoride as a function of pressure is compared with that of nitrogen. Where a maximum occurs in the 60-cycle strength of sulfur hexafluoride, breakdown voltages of five or more times that of nitrogen may be obtained. However, it should be remembered that such breakdown is preceded by violent corona. The corona-onset point in sulfur hexafluoride is of the same order as the breakdown voltage of nitrogen. To take advantage of the high breakdown voltage of sulfur hexafluoride at this pressure, all solid insulation in the field must be capable of withstanding violent corona. Generally, it is more common to design apparatus to operate without the presence of corona. In this case, at a pressure of slightly more than 4 atmospheres, the 60-cycle corona-onset point for sulfur hexafluoride is about $2\frac{1}{2}$ times that of nitrogen. In a uniform field, the sparkover voltage of sulfur hexafluoride is about $2\frac{1}{2}$ times that of nitrogen at 1 atmosphere, and it is much greater than for the divergent fields. Also, it has been noted that with sharper points the ratio of corona onset voltage in sulfur hexafluoride to that in nitrogen is lower. Therefore, nonuniform fields should be avoided, or the divergence of the field limited to the minimum consistent with other design requirements. Where divergent fields exist, the corona-onset voltage is of as great importance as the breakdown voltage. The corona-onset voltage becomes coincident with breakdown for both 60-cycle and positive d-c, beyond a critical pressure which depends on field configuration. For the round-end rod electrode of Fig. 8, this occurs at about 4 atmospheres, as compared with 2 atmospheres for the square-ended rods of Fig. 2. However, negative d-c corona occurs at a somewhat lower voltage for the electrode in Fig. 8, than does the positive. This may be the reason why the 60-cycle sparkover at the higher pressures is lower than that for positive d-c, since the ionization caused by corona occurring on the negative portion of the cycle lowers the breakdown voltage in the same manner as irradiation by X rays.[3]

Impulse Breakdown

The positive impulse-breakdown voltage of sulfur hexafluoride for a 2-inch gap between rods for a $1\frac{1}{2}$x40-microsecond wave is shown in Fig. 9. At about 1-atmosphere pressure, a slight maximum occurs in this curve. This maximum is about half the value of the positive d-c maximum breakdown voltage. When 60-cycle voltage was suddenly applied to this rod gap by a high-voltage switch, the breakdown voltage was not decreased. Therefore, the time required for the space charge around the positive electrode to build up is much less than 1 cycle, but greater than several microseconds, as indicated by the impulse curve. After reaching this maximum, the impulse sparkover voltage decreases only slightly, and then increases with increasing pressure, giving higher values than positive direct current. This seems to indicate that, in the absence of corona, an appreciable time is required for the propagation of positive streamers across the gap to initiate breakdown.

Summary

The occurrence of a maximum in the curve of positive d-c sparkover voltage as a function of pressure as reported in

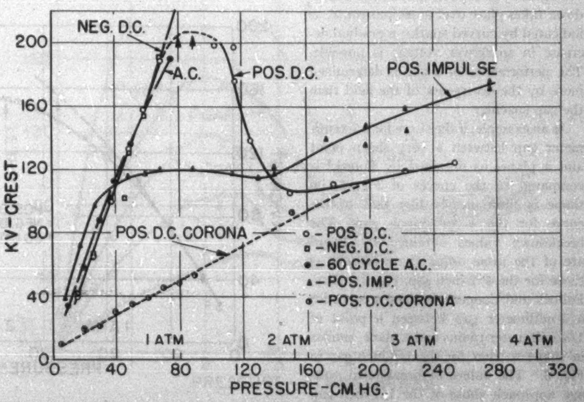


Fig. 9. Sparkover and corona onset voltages of sulfur hexafluoride for a 2-inch gap between 1/2-inch square rods in 12-inch-diameter steel tank, as a function of pressure

the literature was verified for longer gaps. The variation of the 60-cycle sparkover voltage with pressure was found to be very similar to that for positive d-c. The unusually rapid increase in sparkover voltage with increasing electrode spacing was explained by showing the effect of electrode separation on the position of the maximum in the curve of sparkover voltage as a function of pressure. Although positive d-c, negative d-c, and 60-cycle breakdown voltage may increase rapidly with gap spacing over a wide range of pressure, at the higher pressures, the positive d-c and 60-cycle sparkover voltages were found to increase slowly with electrode separation. A situation might well occur in which a considerable increase in electrode separation would not yield any appreciable increase in breakdown voltage. The negative d-c sparkover voltage beyond a critical pressure, which depends on field configuration, was greater than that for positive d-c as predicted by theory. It was shown that, after the maximum was reached, the rate of decrease of breakdown voltage with pressure was a function of field configuration. Although the sharpness of the positive electrode was varied considerably, a pronounced maximum always occurred. This indicated that the same phenomenon would be likely to occur in most electrical apparatus employing an electronegative gas as insulation.

Impulse data showed that the time required for corona stabilization was much less than one cycle of 60-cycle voltage, but more than a few microseconds.

Although breakdown values for sulfur hexafluoride 6 times that of nitrogen have been reported, they occur in the region of a maximum where corona precedes sparkover. To take advantage of this higher breakdown voltage, all solid insulation must be capable of withstanding violent corona.

At the pressure of the maximum, Fig. 8, about 2 atmospheres, the a-c corona onset voltage for sulfur hexafluoride was about 50-per-cent greater than for nitrogen. At 4 atmospheres, the corona-onset voltage for sulfur hexafluoride was about $2^1/_2$ times that of nitrogen. From a practical viewpoint, corona-onset voltage is probably more important to the designer than breakdown voltage.

References

1. Gaseous Insulation for High Voltage Transformers, **G. Camilli, G. S. Gordon, R. E. Plump.** *AIEE Transactions*, vol. 71, pt. III, Jan. 1952, pp. 348–57.

2. Positive Point to Plane Spark Breakdown of Compressed Gases, **T. R. Foord.** *Nature*, London, England, vol. 166, 1950, p. 688.

3. The Effect of Pressure on the Positive Point to Plane Discharge in Nitrogen, Oxygen, Sulfur Hexafluoride, etc., **H. C. Pollock, F. S. Cooper.** *Physical Review*, New York, N. Y., vol. 56, 1939, p. 170.

4. Breakdown Studies in Compressed Gases, **Alvin H. Howell.** *AIEE Transactions (Electrical Engineering)*, vol. 58, 1939, pp. 193–206.

Discussion

T. R. Foord (University of Glasgow, Scotland): No mention is made in the paper concerning the accuracy of the authors' measurements and whether the curves plotted were accurately reproducible. I consider information regarding reproducibility to be particularly important, as it may affect the authors' main conclusion concerning the correlation between the electrode spacing and the maximum in the sparkover-voltage versus gas-pressure curve. In doing similar work, I found that the negative-slope portion of the maxima for some gases was exceedingly difficult, indeed impossible, to reproduce in apparently identical experiments. As much as 80 per cent variation was obtained in extreme cases with Freon *12*. For air, however, when the negative slope is low and extended over a considerable pressure range,[1] good reproducibility (within 5 per cent) was obtained.

The conclusion of the authors that the maximum shifts to a lower pressure for an increased electrode spacing is substantiated by previously published work[1–4] on the positive point-to-plane sparkover in air. It is perhaps interesting to record that a similar effect occurs for a fixed point-to-plane configuration by diluting sulfur hexafluoride with nitrogen. The cause in this case, however, is probably due to a change in the relative predominance of dissociation mechanisms discussed in two recent papers.[5,6] Only by such fundamental investigations, combined with carefully selected and performed sparkover experiments, will a complete understanding of these unusual effects be possible.

References

1. See reference 3 of the paper.

2. See reference 2 of the paper.

3. See reference 4 of the paper.

4. Breakdown of Compressed Gases in Unhomogeneous Fields, **B. M. Hochberg, J. A. Oksman.** *Journal of Technical Physics*, Leningrad, USSR, vol. 5, 1941, p. 39.

5. On Dissociation Processes in Certain Gases of High Dielectric Strength, **J. W. Warren, W. Hopwood, J. D. Craggs.** *Proceedings of the Physical Society*, London, England, vol. 63, 1950, p. 180.

6. Negative Ion Formation in Gases of High Dielectric Strength, **J. W. Warren, J. Marriott, J. D. Craggs.** *Nature*, London, England, vol. 171, 1953, p. 514.

G. Camilli, R. E. Plump, and T. W. Liao (General Electric Company, Pittsfield, Mass.): This paper describes some dielectric peculiarities of SF_6 in a nonuniform field. We can confirm the general observations reported in this paper with some new data of our own. For example, with a 1/4-inch square rod to 3-inch plane, 60-cycle breakdown results were as shown in Fig. 1 of the discussion. In the 1/2-inch gap, breakdown at 45 pounds per square inch gauge pressure was 33 per cent higher than at atmospheric pressure, but in the 1-inch gap the breakdown at 45 pounds per square inch gauge pressure is 13 per cent lower.

In comparison, the results of a 3-inch-diameter round-edge plane-to-plane configuration are shown in Fig. 2 of the discussion and illustrate the classical effect of pressure on the dielectric strength of a gas, and incidentally, are in good agreement with the "capacitor-plate" curve of the paper at atmospheric pressure.

We are also in reasonable agreement with Fig. 7 of the paper which shows maxima in the 1/2-inch rod-to-plane gap. We used a 1/4-inch rod and smaller plane, and found maxima at even lower pressures, that is, at about 100 centimeters of mercury absolute. Our sharper field appears to have brought on the region of anomaly sooner, and has the same effect as the authors' increasing gap distance.

The authors should be commended for their ingenious testing of impressed alternating voltage on the impulse wave, thereby obtaining an idea of the time requirement for corona stabilization. The assumption that the positive half-cycle is most effective in initiating breakdown appears to be well-founded and contributes further to our understanding of electronegative gases.

The fact that strong polarity effects occur in SF_6 should not be surprising. As suggested in other publications,[1,2] there is a variety of energy absorptions which a given gas is capable of, and at breakdown the total effect is a summation of the action and reaction of more than one force. Positive direct current in a nonuniform field diminishes the electronegative nature of SF_6 by sweeping out free electrons, and allows space charge due to positive ions to become much more controlling of the end result.

References

1. See reference 1 of the paper.

2. Fluorine Containing Gaseous Dielectrics, **G. Camilli, R. E. Plump.** *AIEE Transactions*, vol. 72, pt. I, May 1953, pp. 93–102.

C. N. Works and **T. W. Dakin:** The additional data submitted in the discussion by Camilli, Plump, and Liao is very interesting, as it provides an independent confirmation of the breakdown character of this gas. The results obtained with their 1/4-inch rod-to-plane gap supports the statement made in the paper that the performance of the gap is determined more by the divergence of the field than the gap spacing. As they point out, a sharper point causes the maximum to occur at a lower pressure. This is also confirmed by the curves in references 2 and 3 of the paper.

The question raised by Dr. Foord in re-

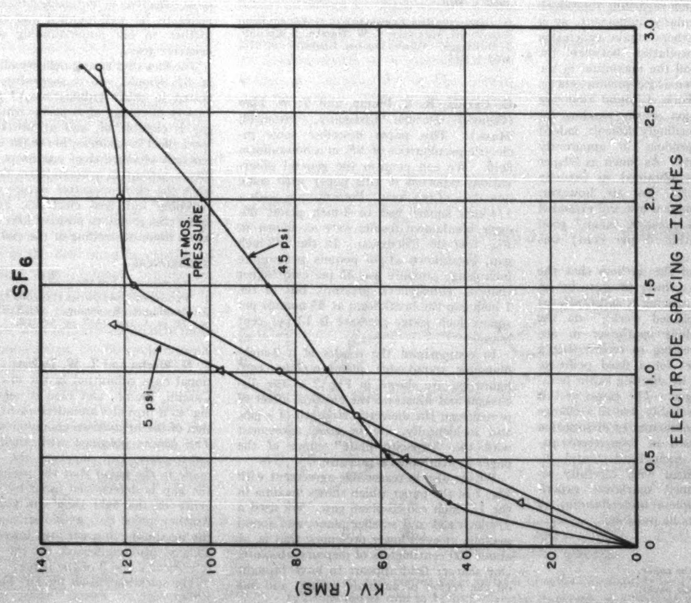


Fig. 1. 60-cycle breakdown voltage versus spacing, 1/4-inch square rod to plane

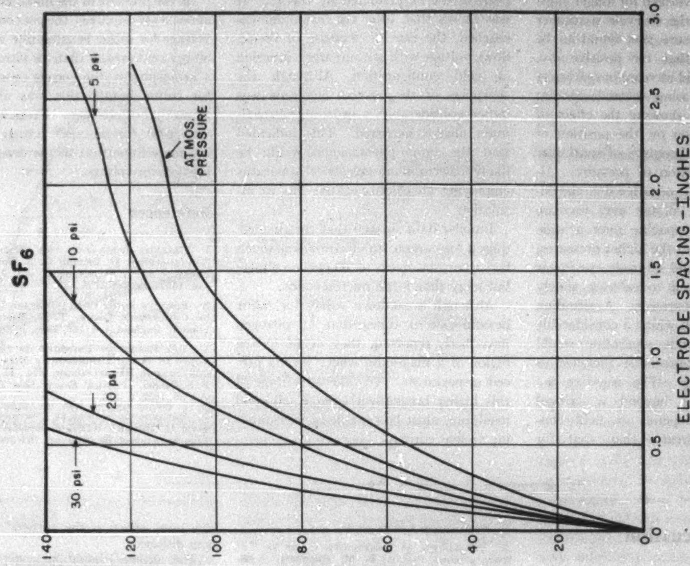


Fig. 2. 60-cycle breakdown voltage versus spacing, 3-inch-diameter round-edged plane-to-plane

gard to the accuracy of the measurements and the reproducibility of the curves is important. The tests were made with well-calibrated testing equipment in all cases, so that the observed breakdown voltages were accurate to about 2 per cent. The scatter of the points from a smooth curve was, with a few exceptions, less than 5 per cent. For the 1/2-inch square rod-to-plane electrodes, the tests which provided the data for some of the curves were repeated and the curves were reproduced to within 10 per cent. All the curves obtained for this electrode configuration were not published, as they are similar to the curves of the paper.

We have done some work with Freon and find that results are much more easily reproducible with SF_6 than with the Freons. Where there is a steep negative slope, as for the 1/2-inch gap in Fig. 3 and the 1-inch gap of Fig. 7 of the paper, there was experimental difficulty. As shown on the curves, either high values or low values of breakdown were obtained and it was not possible to obtain intermediate values. This difficulty only occurred for positive direct current. These steep slopes occur only for certain spacings for the various electrode arrangements. A possible explanation of this steep slope is given in reference 3 of the paper. These steep slopes probably indicate an unstable condition, but it is definite in these cases that there is an abrupt change in breakdown voltage at a critical pressure.

It is probable that decomposition-product molecules and ions participate in the corona-discharge process where the corona persists. This would likely cause fluctuations in the results, which might be greater in the case of more complex molecules. In the case of more complex molecules, the product molecules formed by recombination of fragments produced by the electron-initiated dissociation are appreciably different from the parent molecule. Unless the concentration of these decomposition products has reached an equilibrium value, where their rate of formation and diffusion away are balanced, the breakdown voltage observed following corona might fluctuate depending on the transient concentration of these products.

In all cases the values of the maxima were accurately established, and are believed to be reproducible with reasonable accuracy. In support of this statement, it is pointed out that data in this paper check closely data in an earlier paper by Camilli and Chapman,[1] and that the new data by Camilli, Plumb, and Liao are in good agreement with the data presented in this paper.

We agree with Dr. Foord that fundamental studies of the dissociation mechanisms of SF_6 should be combined with sparkover experiments to enable us to gain a better understanding of the breakdown mechanisms that may occur in this gas.

Reference

1. Gaseous Insulation for High-Voltage Apparatus, G. Camilli, J. J. Chapman. *AIEE Transactions* vol. 66, 1947, pp. 1463–70.

Voltage Regulation of 12-Phase Double-Way Rectifiers

R. L. WITZKE
FELLOW AIEE

J. V. KRESSER
MEMBER AIEE

J. K. DILLARD
MEMBER AIEE

Summary

Mathematical expressions for calculating the voltage regulation of double-way rectifiers connected for 12-phase operation have been derived and tabulated for ready reference. Calculated curves are presented for estimating voltage regulation, angle of delay, and overlap angle for any load out to short circuit. Regulation curves taken in the laboratory are included to substantiate the general trend of the calculated curves.

This paper is an extension of the 6-phase case;[1] the same notation is used, and similar treatment of the subject is employed in order to facilitate comparison.

CONVENTIONAL METHODS of calculating rectifier voltage regulation are valid only if the angle of overlap is less than 360 degrees divided by the number of phases. At larger angles of overlap an inherent delay angle, introduced by a-c reactance common to two or more simple rectifiers, reduces the direct voltage and thereby increases the voltage regulation. This phenomenon has been described in two papers,[1,2] recently presented to the Institute, in which curves and equations for determining voltage regulation of 6-phase rectifiers were given. Two earlier articles[3,4] recognized the phenomenon in 6-phase circuits.

Twelve-phase rectification obtained by connecting two double-way rectifiers in series on the d-c side is often employed in special applications requiring high direct voltages. The customary practice is to supply the rectifiers from two 3-phase transformer banks in parallel on the a-c side, one bank connected wye-delta, the other delta-delta. Operation of such rectifiers is affected to a greater extent by common a-c reactance than is operation of 6-phase rectifiers. This paper considers this 12-phase operation.

Nomenclature

α = first angle of forced commutation delay
α_m = maximum value of α
β = second angle of delay
$\gamma = u - 60°$ = interval of simultaneous commutations within one bridge
$\delta = u - 30°$ = interval of simultaneous commutations between two bridges
$\omega = 2\pi f$ = angular frequency
$E_{a,b,c}$ = source line-to-neutral instantaneous voltages
$E_{a1,b1,c1}$ = line-to-neutral voltages on primary of bridge I transformer
$E_{a2,b2,c2}$ = line-to-neutral voltages on primary of bridge II transformer
$E_{1,3,5}$ = line-to-neutral voltages on secondary of bridge I transformer
$E_{2,4,6}$ = line-to-neutral voltages on secondary of bridge II transformer
E_{a1}' = instantaneous value of E_{a1} during commutation in the other bridge
e_d = instantaneous direct voltage
E_d = average direct voltage under load
E_{d0} = average direct voltage at no load
E_s = transformer secondary line-to-neutral rms voltage at no load
f = frequency of system
$i_{1,2,3,4,5}$ = instantaneous value of commutation currents on primary side of transformers
$i_{1c,2c,3c,4c,5c}$ = instantaneous value of commutation currents through rectifier tubes
I_c = commutated direct current
I_d = load current
I_dX_c/E_s = reactance factor based on load current
$k = X_s/X_c$ = ratio of common to total commutating reactance
L_L = inductance not common to both bridges
L_s = inductance common to both bridges
$L = L_s + L_L$ = commutating inductance
$p = d/dt$
t = time in seconds
u = commutating (overlap) angle
X_L = reactance not common to both bridges
X_s = reactance common to both bridges
$X_c = X_s + X_L$ = commutating reactance

Paper 53-242, recommended by the AIEE Electronic Power Converters Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 17, 1953; made available for printing April 16, 1953.

R. L. Witzke, J. V. Kresser, and J. K. Dillard are with the Westinghouse Electric Corporation, in McKeesport, Pa., San Francisco, Calif., and East Pittsburgh, Pa., respectively.

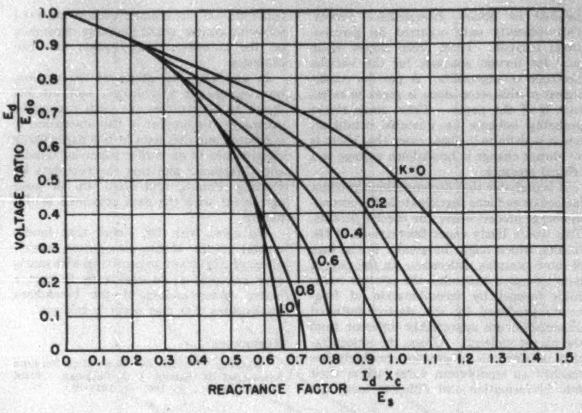


Fig. 1. Voltage regulation curves for 12-phase double-way rectifier

Table I. Limiting Values of Angles in the Several Modes of Operation

Mode	u	α	β
1	$0 \rightarrow 30°$	0	0
2	$30°$	$0 \rightarrow \alpha_m$	0
3	$30° \rightarrow 60°$	α_m	0
4	$60°$	α_m	$0 \rightarrow 30°$
5	$60° \rightarrow (120° - 2\alpha_m)$	α_m	$30°$

The chief difference between the 12-phase and 6-phase cases is the larger number of modes of operation in the 12-phase rectifier as load is increased from zero to short-circuit value. There are five modes of operation instead of three (instead of four, where leakage reactance between secondary windings is not negligible[5]). In the first mode of operation, as load current is increased from no load, the commutating or overlap angle u increases from zero to 30 degrees. During the second mode of operation, the commutating angle is fixed at 30 degrees, but the start of commutation is delayed by an inherent delay angle α which varies from zero to some maximum value α_m as the rectifier supplies increasing load current. Operation in the second mode is quite similar to 6-phase operation where leakage reactance between secondary windings is not negligible.[5] As is shown in the Appendix

$$\tan \alpha_m = \frac{k\sqrt{3}}{4-3k}$$

where k is the ratio of reactance common to both 6-phase rectifiers to total commutating reactance. The angle of inherent delay does not change during the third mode, while the commutating angle increases from 30 to 60 degrees. During the fourth mode of operation, the commutating angle is fixed at 60 degrees, and the start of commutation is delayed by a second angle of delay β, which increases from zero to 30 degrees. In the fifth mode, the total angle of delay $(\alpha_m+\beta)$ is fixed, while the commutating angle increases from 60 to $(120-2\alpha_m)$ degrees at short circuit. The five modes of operation are summarized in Table I.

The ratio of direct voltage to no-load direct voltage (E_d/E_{do}) is plotted against reactance factors in Fig. 1. These curves show that voltage regulation increases as the ratio of common to total reactance becomes larger.

Calculated values of commutating angle u are plotted against reactance factors in Fig. 2. To see the five modes of operation, consider $k=0.4$. The commutating angle increases from zero to 30 degrees at $I_dX_c/E_s=0.164$ during the first mode. The angle is then fixed at 30 degrees until the end of the second mode at $I_dX_c/E_s=0.307$. There is an increase in the angle from 30 to 60 degrees during the third mode, which ends at $I_dX_c/E_s=0.632$. Then the angle is fixed until the end of the fourth mode at $I_dX_c/E_s=0.875$. In the fifth mode, the angle increases to 92 degrees at $I_dX_c/E_s=0.975$, the short-circuit value.

Calculated values of angle of delay are plotted against reactance factors in Fig. 3. Both the first and second angle of inherent delay are given; for example, for $k=0.4$, the first angle α increases from zero to 14 degrees in the interval of I_dX_c/E_s from 0.164 to 0.307 and the second angle β increases from zero to 30 degrees, making a change from 14 to 44 degrees in the total angle of delay, in the interval of I_dX_c/E_s from 0.632 to 0.875.

Equations used in plotting the curves are derived in the Appendix.

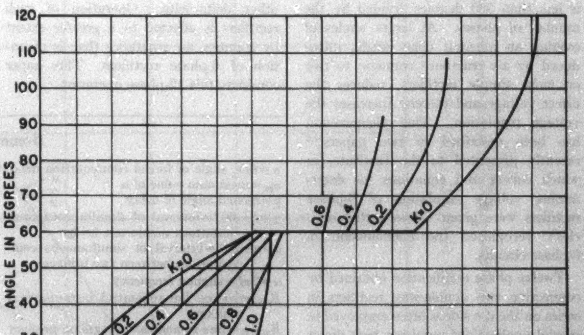


Fig. 2. Commutation angle as a function of reactance factor

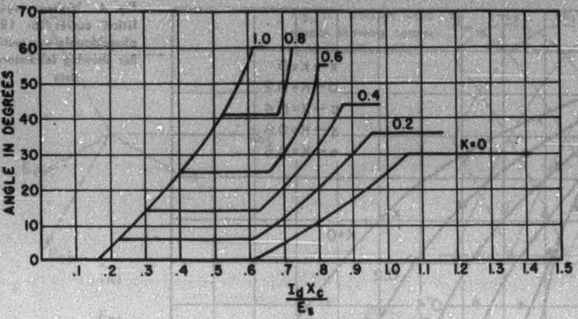


Fig. 3. Angles of delay as a function of reactance factor

Summary of Equations

Mode 1

$$\frac{I_d X_c}{E_s} = \frac{\sqrt{6}}{2}(1-\cos u)$$

$$\frac{E_d}{E_{do}} = \frac{1}{2}(1+\cos u)$$

Mode 2

$$\frac{I_d X_c}{E_s} = \sqrt{6}\sin 15^\circ \sin(\alpha+15^\circ)$$

$$\frac{E_d}{E_{do}} = \cos 15^\circ \cos(\alpha+15^\circ)$$

Mode 3

$$\frac{I_d X_c}{E_s} = \left[\frac{3+\sqrt{3}}{2}\left(\frac{2-k\sqrt{3}}{2+k\sqrt{3}}\right)\sin\left(\frac{u}{2}-15^\circ\right)+ \frac{3-\sqrt{3}}{2}\cos\left(\frac{u}{2}-15^\circ\right)\right]\sin\left(\alpha_m+\frac{u}{2}\right)$$

$$\frac{E_d}{E_{do}} = \frac{1-\sqrt{3}}{2\sqrt{2}}\left[\frac{2+k\sqrt{3}}{2-k\sqrt{3}}\sin\left(\frac{u}{2}-15^\circ\right)- (2+\sqrt{3})\cos\left(\frac{u}{2}-15^\circ\right)\right]\cos\left(\alpha_m+\frac{u}{2}\right)$$

Mode 4

$$\frac{I_d X_c}{E_s} = \frac{\sqrt{6}}{2+k\sqrt{3}}\sin(\alpha_m+\beta+30^\circ)$$

$$\frac{E_d}{E_{do}} = \frac{\sqrt{3}(1-k)}{2-k\sqrt{3}}\cos(\alpha_m+\beta+30^\circ)$$

Mode 5

$$\frac{I_d X_c}{E_s} = \frac{\sqrt{2}}{(1+k)(2+k\sqrt{3})}\times \left[\sin\left(\frac{u}{2}-30^\circ\right)+\sqrt{3}(1+k)\times \cos\left(\frac{u}{2}-30^\circ\right)\right]\sin\left(\frac{u}{2}+\alpha_m+30^\circ\right)$$

$$\frac{E_d}{E_{do}} = \frac{\sqrt{3}(1-k)}{2-k\sqrt{3}}\left[\cos\left(\frac{u}{2}-30^\circ\right)- \sqrt{3}(1+k)\sin\left(\frac{u}{2}-30\right)\right]\cos\left(\frac{u}{2}+\alpha_m+30^\circ\right)$$

Laboratory Tests

The trend of the calculated regulation curves was substantiated by tests on a model rectifier on the analogue computer, or Anacom. The technique of handling such problems on the Anacom was described in the paper on 6-phase rectifier regulation.[1] As in that case, the rectifier elements were represented by selected selenium cells having low voltage drop in the forward direction and very high impedance in the reverse direction. The circuit was supplied from a 19-volt 3-phase source of low impedance, providing an essentially infinite bus.

Data taken on the Anacom circuit are shown in Fig. 4. These data have been corrected for voltage drop across the rectifying elements and for losses in the circuit elements. The solid lines are the calculated curves taken from Fig. 1. Although the trend of the calculated curves were substantiated by Anacom data, there is some discrepancy between the laboratory data and the theoretical curves at the higher ratios of common to total reactance.

One handicap, peculiar to the 12-phase study, lay in the construction of the Anacom transformers. These transformers are provided with a variable turns ratio in 10-per-cent steps. Therefor, the delta-delta transformer bank could be set up with a 100-per-cent turns ratio, but the turns ratio of the wye-delta bank had to be set up with a 60-per-cent turns ratio instead of the desired 57.7 per cent. Some inaccuracies were introduced into the laboratory curves by the method employed to account for losses in the computer circuit elements. Data taken from the Anacom were first corrected for the tube drop corresponding to the load current. Then corrections were made for resistance drop by dividing the watts loss in the a-c circuit by the current in the d-c circuit, the conventional procedure of correcting for a-c circuit losses in rectifiers. The accuracy of this method depends upon how closely the current wave forms approach theoretical shapes. These two factors, lack of the exact transformer turns ratio required, and variation of the current wave from the exact theoretical shape, account for the differences between calculated curves and laboratory data.

Appendix. Twelve-Phase Operation With Two Double-Way Rectifiers in Series

A circuit consisting of two 6-phase full-wave bridges with d-c outputs connected in series is shown in Fig. 5. Transformer turns ratios are selected to give unity ratio of transformation between line-to-line voltages. As shown on the phasor diagram of Fig. 6, line-to-neutral voltages behind source reactance are

$$E_a = \sqrt{2}E_s\cos\left(\omega t+\frac{\pi}{2}\right)$$

$$E_b = \sqrt{2}E_s\cos\left(\omega t-\frac{\pi}{6}\right) \quad (1)$$

$$E_c = \sqrt{2}E_s\cos\left(\omega t-\frac{5\pi}{6}\right)$$

where E_s is the rms value of transformer-secondary no-load line-to-neutral voltage.

At no load, the voltages applied to the anodes of one 3-phase rectifier group and to the cathodes of the other 3-phase rectifier group of each bridge are:

Wye-Delta (bridge I)

$$E_1 = \sqrt{2}E_s\cos\left(\omega t-\frac{\pi}{3}\right)$$

$$E_3 = \sqrt{2}E_s\cos(\omega t-\pi)$$

$$E_5 = \sqrt{2}E_s\cos\left(\omega t+\frac{\pi}{3}\right)$$

Delta-Delta (bridge II) (2)

$$E_2 = \sqrt{2}E_s\cos\left(\omega t-\frac{5\pi}{6}\right)$$

$$E_4 = \sqrt{2}E_s\cos\left(\omega t+\frac{\pi}{2}\right)$$

$$E_6 = \sqrt{2}E_s\cos\left(\omega t-\frac{\pi}{6}\right)$$

These voltages determine the normal time of commutation of the first 3-phase rectifier group of each bridge, and the negatives of these voltages determine the normal time of commutation of the second 3-phase rectifier group of each bridge. The 30-degree phase shift between the two bridges results in equivalent 12-phase operation.

Referring to Fig. 5, the reactance $X_L = 2\pi f L_L$ includes the transformer reactance and all reactance between transformer primary terminals and the common point of

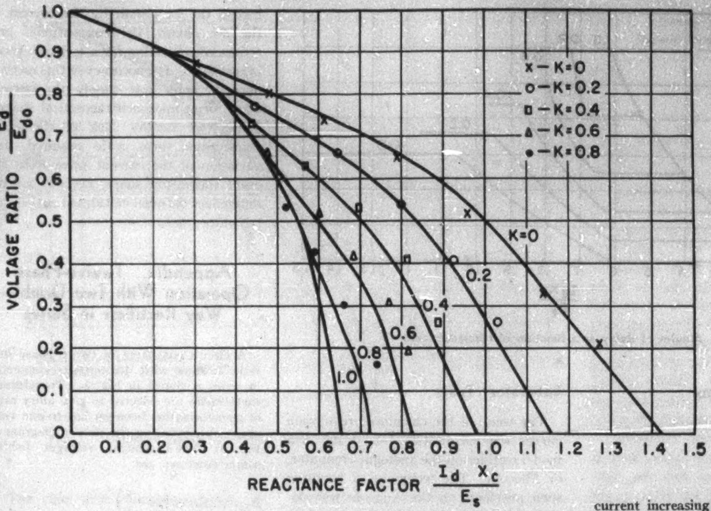


Fig. 4. Voltage regulation curves for 12-phase double-way rectifier showing laboratory data

the supply circuit. The reactance X_s includes all reactance between this common point and the source.

First Mode of Operation

During the first mode of operation, commutations occur at different times in the two bridges, that is, a commutation in bridge I is completed before the start of a commutation in bridge II.

Referring to Fig. 5, the tubes are numbered in the order in which they start to conduct and the current i_1 represents an instantaneous current increasing in T_1 and decreasing in T_9 during the commutation from T_9 to T_1. The following are obtained

$$E_{ab} = -\frac{3}{2}Lpi_1 + E_{a1b1} \tag{3}$$

$$E_{bc} = E_{b1c1}$$

Since a T_9 to T_1 commutation imposes a temporary short circuit on terminals 5 to 1 on the secondary of the bridge I transformer, $E_{a1}=0$; and $E_{a1b1}=E_{c1a1}=-E_{b1c1}/2$. Therefore, from equations 3

$$pi_1 = -\frac{2E_{ab}+E_{bc}}{3L} = -\frac{E_a}{L} \tag{4}$$

The current i_2 represents an instantaneous current increasing in T_2 and decreasing in T_{10} during the commutation from T_{10} to T_2. During this commutation, the following equation can be written

$$E_{ca} = 2Lpi_2 + E_{c2a2} \tag{5}$$

Since a T_{10} to T_2 commutation imposes a temporary short circuit on terminals 4 to 2 on the secondary of the bridge II transformer, $E_{c2a2}=0$, and

$$pi_2 = \frac{E_{ca}}{2L} \tag{6}$$

During the first mode of operation, the average direct voltage can be obtained by taking the average of the instantaneous

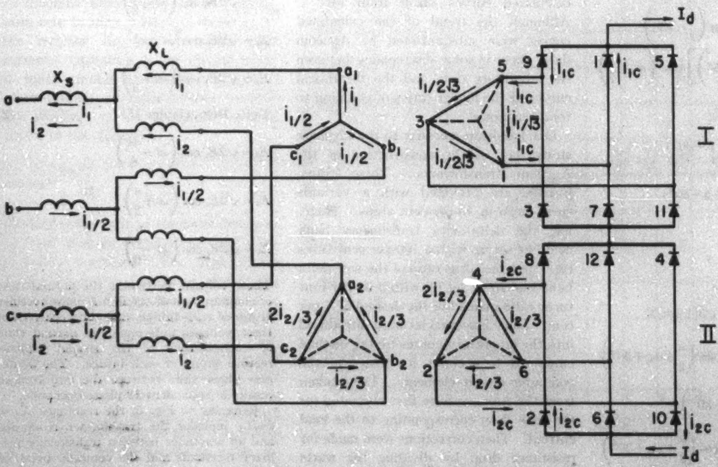


Fig. 5. Circuit diagram showing currents for commutations from T_9 to T_1 and T_{10} to T_2

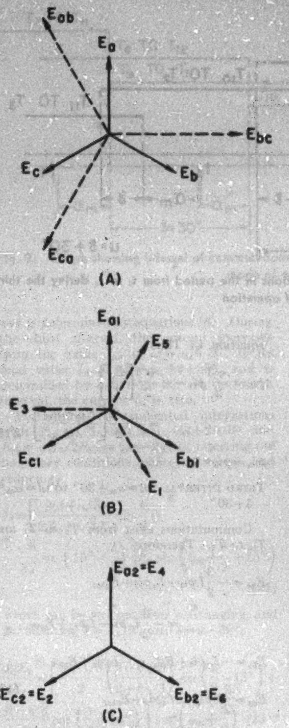


Fig. 6. Phasor diagrams

A. Line-to-neutral voltage behind source reactance
B. No-load voltages on wye-delta bank
C. No-load voltages on delta-delta bank

values over a 60-degree period, since there is a commutation every 60 degrees in each bridge. The average direct voltages of the two bridges are equal, the instantaneous values being displaced by 30 degrees. Therefore, the average direct voltage for two bridges in series is twice the average voltage of a single bridge,[1] or

$$E_d = \frac{6\sqrt{6}}{\pi} E_s \left(\frac{1+\cos u}{2} \right) \tag{7}$$

Dividing equation 7 by the no-load direct voltage, $E_{do} = 6\frac{\sqrt{6}}{\pi} E_s$

$$\frac{E_d}{E_{do}} = \frac{1+\cos u}{2} \tag{8*}$$

The commutating angle is the same as for a single bridge,[1] or

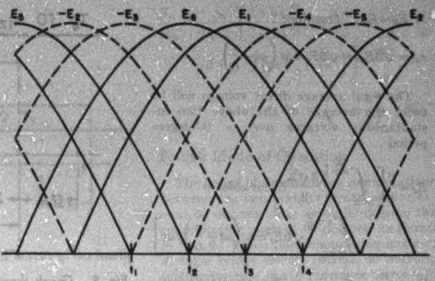
$$\cos u = 1 - \frac{2I_d X_c}{\sqrt{6}E_s}$$

from which

$$\frac{I_d X_c}{E_s} = \frac{\sqrt{6}}{2}(1-\cos u) \tag{9*}$$

The limitation on u is 30 degrees, rather than 60 degrees as for a single bridge. That is, when the reactance factor is greater than the value for which $u=30$ degrees, forced delay is introduced, and the rectifier enters the second mode of operation.

Fig. 7. Voltage wave shapes showing normal commutation times



Second Mode of Operation

During the second mode of operation the angle of overlap u remains fixed at 30 degrees and the start of commutations in each bridge is delayed by an angle α, which depends on the ratio of reactances X_L and X_s. As in the first mode of operation, however, a commutation in bridge I is completed before the start of the subsequent commutation in bridge II.

Normal start of commutation from T_9 to T_1 on bridge I, designated by t_1 on Fig. 7, is when $E_5 = E_1$, that is, when $E_{a1} = E_{b1}/\sqrt{3} = 0$. If at that time the commutation from T_8 to T_{12} is still going on in bridge II, equations similar to 5 and 6 hold for the T_8 to T_{12} commutation. During this commutation

$$pi_2 = -\frac{E_{ab}}{2L}$$

Note that $i_1=0$ until the T_9 to T_1 commutation starts. Therefore, following Fig. 5, $E_{a1}' = E_a + L_s pi_2$ or

$$E_{a1}' = E_a - \frac{L_s}{2L} E_{ab} = \frac{(2X_c - X_s)E_a + X_s E_b}{2X_c} \tag{10}$$

The normal start of the T_9 to T_1 commutation has been selected at $\omega t=0$. Taking E_a and E_b from equation 1 at this instant, $E_{a1}'>0$. Therefore, for reactance factors greater than those for which $u=30$ degrees, forced delay in the start of the T_9 to T_1 commutation is introduced. By the same reasoning, the start of the T_8 to T_{12} commutation was delayed while T_7 to T_{11} were still commutating. This is similar to the single bridge circuit,[1] except that forced delay is introduced at $u=30$ degrees instead of $u=60$ degrees. Another difference is that the maximum delay angle is independent of the reactance factor, and is a function of the ratio of reactances only, as will be shown.

At some angle α_m, which is the angle of maximum forced delay, E_{a1}' will become zero. Letting $\omega t = \alpha_m$ in equation 1, substituting these values in equation 10, and equating to zero

$$E_{a1}' = \frac{(2X_c - X_s)}{2X_c}\sqrt{2}E_s \cos\left(\alpha_m + \frac{\pi}{2}\right) + \frac{X_s}{2X_c}\sqrt{2}E_s \cos\left(\alpha_m + \frac{\pi}{6}\right) = 0$$

Carrying out the trigonometry

$$\tan \alpha_m = \frac{X_s\sqrt{3}}{4X_c - 3X_s} = \frac{k\sqrt{3}}{4-3k} \tag{11*}$$

Since the largest value of k is unity ($X_s = X_c$ where $X_L=0$), the largest maximum delay angle is 60 degrees, as compared with 30 degrees for a single bridge circuit.[1] Also, from similarity with the single bridge, the commutating angle remains fixed at 30 degrees (rather than 60 degrees) for reactance factors greater than $(\sqrt{6}/2)\times(1-\cos 30°)$.

The average direct voltage is obtained from the instantaneous values over a 60-degree period. For bridge I, starting from the instant that the T_9 to T_1 commutation starts, in the interval $\omega t = \alpha$ to $\omega t = \alpha + 30°$, T_{11}, T_9, and T_1 are conducting.

Therefore

$$e_{d1} = -E_3 + E_1 = \sqrt{3}E_{b1}$$

Since there is no commutation taking place in bridge II during this interval, $E_{b1} = -E_{c1}$, and from equations 3

$$E_{bc} = E_{b1c1} = 2E_{b1}$$

Therefore

$$e_{d1} = \frac{\sqrt{3}}{2}E_{bc} = \frac{3\sqrt{2}}{2}E_s \cos \omega t$$

In the interval $\omega t = \alpha + 30°$ to $\omega t = \alpha + 60°$, the T_9 to T_1 commutation being completed, T_{11} and T_1 are conducting in bridge I. There is a commutation going on in bridge II, the T_{10} to T_2 commutation starting 30 degrees after the start of the T_9 to T_1 commutation. From Fig. 5

$$E_{ab} = -L_s pi_1 + E_{a1b1}$$

$$E_{ac} = -2L_s pi_2 + E_{a1c1}$$

which yield, upon simplifying

$$2E_{ab} - E_{ac} = 2E_{a1b1} - E_{a1c1}$$

---

* Equations marked * are tabulated in the Summary section.

or $E_b = E_{b1}$. Therefore

$$e_{d2} = \sqrt{3}E_{b1} = \sqrt{6}E_s \cos\left(\omega t - \frac{\pi}{6}\right)$$

The total average direct voltage will be twice the average of the above two instantaneous voltages over a 60-degree period

$$E_d = \frac{6}{\pi}\left[\int_{\alpha}^{\alpha+\frac{\pi}{6}} \frac{3\sqrt{2}}{2}E_s \cos \omega t d\omega t + \int_{\alpha+\frac{\pi}{6}}^{\alpha+\frac{\pi}{3}} \sqrt{6}E_s \cos\left(\omega t - \frac{\pi}{6}\right) d\omega t\right]$$

Evaluating the integrals between the indicated limits, expanding, and collecting terms

$$E_d = \frac{6\sqrt{6}}{\pi}E_s \cos 15° \cos(\alpha + 15°)$$

But

$$E_{do} = \frac{6\sqrt{6}}{\pi}E_s$$

giving

$$\frac{E_d}{E_{do}} = \cos 15° \cos(\alpha + 15°) \qquad (12)^*$$

The commutating current, from Fig. 5, is $i_{1c} = (\sqrt{3}/2)i_1$ and $i_{1c} = I_d$ when $\omega t = \alpha + 30°$. Therefore, from equation 4

$$I_d = \int_{\alpha}^{\alpha+\frac{\pi}{6}} \frac{\sqrt{6}E_s}{2X_c} \sin \omega t d\omega t \qquad (13)$$

Integrating and simplifying

$$\frac{I_d X_c}{E_s} = \sqrt{6} \sin 15° \sin(\alpha + 15°) \qquad (14)^*$$

Limits on the second mode of operation are $\alpha = 0$ to $\alpha = \alpha_m$ where α_m is obtained from equation 11. Thus

$$\frac{\sqrt{6}}{2}(1 - \cos 30°) \leqq \frac{I_d X_c}{E_s} \leqq \frac{3-\sqrt{3}}{2} \sin(\alpha_m + 15°) \qquad (15)$$

Third Mode of Operation

For reactance factors greater than those for which $\alpha = \alpha_m$ in equation 14, commutation will start at α_m since at that point the two instantaneous voltages E_6 and E_1 become equal. Hence, there will be portions during the commutation period where both bridges will commutate simultaneously. During the third mode of operation the commutation (overlap) angle will increase with the increase in the reactance factor. This mode of operation continues until $u = 60°$.

Because of the symmetry of operation the simultaneous commutation interval δ is the same at the end as at the start of the commutating period, as shown in Fig. 8. Since a new commutation starts every 30°, the total commutating period $u = 30° + \delta$. However, in each bridge commutations start 60 degrees apart, making the interval of symmetry 60 degrees. Thus $u \leqq 60°$ and $\delta \leqq 30°$ during the third mode of operation.

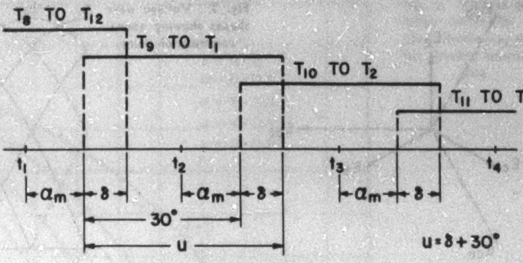


Fig. 8. Graph showing interval of commutations in the period from t_1 to t_4 during the third mode of operation

The 60-degree angle of symmetry is composed of four distinct intervals as shown in Fig. 8. During the first interval, $\omega t = \alpha_m$ to $\omega t = \alpha_m + \delta$, simultaneous commutations exist from T_9 to T_1 (in bridge I) and T_8 to T_{12} (in bridge II). In the second interval, $\omega t = \alpha_m + \delta$ to $\omega t = \alpha_m + 30°$, the only commutation is T_9 to T_1. During the third interval, $\omega t = \alpha_m + 30°$ to $\omega t = \alpha_m + 30° + \delta$, simultaneous commutations exist from T_9 to T_1 (in bridge I) and T_{10} to T_2 (in bridge II). In the fourth interval, $\omega t = \alpha_m + 30° + \delta$ to $\omega t = \alpha_m + 60°$, the only commutation is T_{10} to T_2.

Referring to Fig. 5, the current i_2 is shown for the T_{10} to T_2 commutation. For the T_8 to T_{12} commutation, a similar current i_4 (not shown) flows in terminal b and out terminal a of Fig. 5. Bearing this fact in mind, the following equations apply for the 60-degree period, beginning at the start of the T_9 to T_1 commutation:

FIRST INTERVAL, $\omega t = \alpha_m$ to $\omega t = \alpha_m + \delta$

Commutations exist from T_9 to T_1 and T_8 to T_{12}. Therefore

$$\begin{aligned}
E_{ab} &= -\frac{3}{2}Lpi_1 - 2L_s pi_4 + E_{a1b1} \\
&= -\frac{3}{2}L_s pi_1 - 2Lpi_4 + E_{a2b2} \\
E_{bc} &= L_s pi_4 + E_{b1c1} = Lpi_4 + E_{b2c2} \\
E_{ca} &= \frac{3}{2}Lpi_1 + L_s pi_4 + E_{c1a1} \\
&= \frac{3}{2}L_s pi_1 + Lpi_4 + E_{c2a2}
\end{aligned} \qquad (16)$$

$E_{a1} = 0$ and $E_{a2b2} = 0$

As shown in Fig. 5, $i_{1c} = (\sqrt{3}/2)i_1$ and $i_{4c} = i_4$. Therefore, after solving equations 16

$$\begin{aligned}
pi_{1c} &= \frac{\sqrt{3}}{2}pi_1 \\
&= -\frac{1}{\sqrt{3}}\left[\frac{(2L - 3L_s)E_{ab} - 2LE_{ca}}{(4L^2 - 3L_s^2)}\right] \\
pi_{4c} &= pi_4 = -\left[\frac{(2L - L_s)E_{ab} + L_s E_{ca}}{(4L^2 - 3L_s^2)}\right]
\end{aligned} \qquad (17)$$

SECOND INTERVAL, $\omega t = \alpha_m + \delta$ to $\omega t = \alpha_m + 30°$

The only commutation is T_9 to T_1, and the current has already been derived in equation 4. Thus

$$\begin{aligned}
pi_{1c} &= \frac{\sqrt{3}}{2}pi_1 = -\frac{\sqrt{3}}{2}\left[\frac{E_a}{L}\right] \\
&= -\frac{\sqrt{3}}{2}\left[\frac{E_{ab} - E_{ca}}{3L}\right]
\end{aligned} \qquad (18)$$

$pi_{2c} = pi_2 = 0$

THIRD INTERVAL, $\omega t = \alpha_m + 30°$ to $\omega t = \alpha_m + \delta + 30°$

Commutations exist from T_9 to T_1 and T_{10} to T_2. Therefore

$$\begin{aligned}
E_{ab} &= -\frac{3}{2}Lpi_1 - L_s pi_2 + E_{a1b1} \\
&= -\frac{3}{2}L_s pi_1 - Lpi_2 + E_{a2b2} \\
E_{bc} &= -L_s pi_2 + E_{b1c1} = -Lpi_2 + E_{b2c2} \\
E_{ca} &= \frac{3}{2}Lpi_1 + 2L_s pi_2 + E_{c1a1} \\
&= \frac{3}{2}L_s pi_1 + 2Lpi_2 + E_{c2a2}
\end{aligned} \qquad (19)$$

$E_{a1} = 0$ and $E_{c2a2} = 0$

After solving equations 19

$$\begin{aligned}
pi_{1c} &= \frac{\sqrt{3}}{2}pi_1 \\
&= -\frac{1}{\sqrt{3}}\left[\frac{2LE_{ab} - (2L - 3L_s)E_{ca}}{(4L^2 - 3L_s^2)}\right] \\
pi_{2c} &= pi_2 = \frac{L_s E_{ab} + (2L - L_s)E_{ca}}{(4L^2 - 3L_s^2)}
\end{aligned} \qquad (20)$$

FOURTH INTERVAL, $\omega t = \alpha_m + \delta + 30°$ to $\omega t = \alpha_m + 60°$

In this interval the only commutation is T_{10} to T_2 and the current has already been derived in equation 6

$$\begin{aligned}
pi_{1c} &= \frac{\sqrt{3}}{2}pi_1 = 0 \\
pi_{2c} &= pi_2 = \frac{E_{ca}}{2L}
\end{aligned} \qquad (21)$$

During the first interval the current i_{1c} increases from zero at $\omega t = \alpha_m$ to some value I_{d1} at $\omega t = \alpha_m + \delta$, and is determined by equations 17. In the second interval, the current increases from the value I_{d1} at $\omega t = \alpha_m + \delta$ to some value I_{d2} at $\omega t = \alpha_m + 30°$,

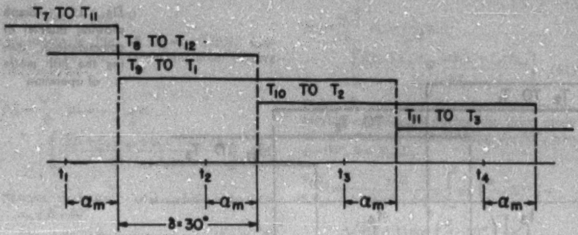


Fig. 9. Graph showing interval of commutations in the period from t_1 to t_4 at start of the fourth mode of operation

and is determined by equations 18. During the third interval, the current increases from the value I_{d2} at $\omega t=\alpha_m+30°$ to its final value I_d at $\omega t=\alpha_m+\delta+30°$, and is determined by equations 20. In the fourth interval, the current i_{1c} is zero.

Performing the indicated integrations with $E_{ab}=\sqrt{3}\sqrt{2}E_s \cos(\omega t+2\pi/3)$ and $E_{ca}=\sqrt{3}\sqrt{2}E_s \cos(\omega t-2\pi/3)$, inserting the boundary conditions, collecting terms, and reducing

$$I_d=\left[\frac{2(3+\sqrt{3})}{2X_c+\sqrt{3}X_s}\sin\frac{\delta}{2}+\frac{\sqrt{6}}{X_c}\sin\left(15°-\frac{\delta}{2}\right)\right]E_s\sin\left(\alpha_m+\frac{\delta}{2}+15°\right) \tag{22}$$

which can be written after rearranging, and substituting $k=X_s/X_c$ and $\delta=u-30°$.

$$\frac{I_dX_c}{E_s}=\left[\frac{3+\sqrt{3}}{2}\left(\frac{2-k\sqrt{3}}{2+k\sqrt{3}}\right)\sin\left(\frac{u}{2}-15°\right)+\frac{3-\sqrt{3}}{2}\cos\left(\frac{u}{2}-15°\right)\right]\sin\left(\alpha_m+\frac{u}{2}\right) \tag*{(23)*}$$

As before, the average direct voltage is obtained from the component instantaneous voltages averaged over a 60-degree period. Taking bridge I, and referring to Fig. 6(B)

$$e_d=-E_3+E_1=\sqrt{3}E_{b1} \tag{24}$$

where E_{b1} is determined for the particular interval considered.

In the first interval, from equations 16, where $E_{a1}=0$

$$e_{d1}=-\sqrt{3}\times\left[\frac{(2X_c^2-X_cX_s-X_s^2)E_{ab}+2(X_c^2-X_s^2)E_{ca}}{4X_c^2-3X_s^2}\right] \tag{25}$$

In the second interval, the only commutation is T_9 to T_1. Therefore, from equations 3

$$e_{d2}=\sqrt{3}E_{b1}=\sqrt{3}\left(-E_{ab}-\frac{3}{2}Lpi_1\right)$$

Substituting equation 18 and reducing

$$e_{d2}=\frac{\sqrt{3}}{2}E_{bc} \tag{26}$$

In the third interval from equations 19

$$e_{d3}=\sqrt{3}E_{b1}=\sqrt{3}\left(-E_{ab}-\frac{3}{2}Lpi_1-L_spi_2\right)$$

Substituting equations 20

$$e_{d3}=-\sqrt{3}\left[\frac{2(X_c^2-X_s^2)E_{ab}+(2X_c^2-X_cX_s-X_s^2)E_{ca}}{4X_c^2-3X_s^2}\right] \tag{27}$$

For the fourth interval, the only commutation is T_{10} to T_2, and in bridge I, T_{11} and T_1 are conducting. From Fig. 5

$$E_{ab}=-L_spi_2+E_{a1b1}$$

$$E_{ca}=2L_spi_2+E_{c1a1}$$

or

$$2E_{ab}+E_{ca}=2E_{a1b1}+E_{c1a1}=-3E_{b1}$$

from which

$$e_{d4}=\sqrt{3}E_{b1}=-\left[\frac{2E_{ab}+E_{ca}}{\sqrt{3}}\right] \tag{28}$$

Integrating each voltage of equations 25, 26, 27, and 28 between the limits of its interval as given above and dividing by $\pi/3$ to obtain the average voltage, reducing, and further dividing by $E_{do}=(3\sqrt{6}/\pi)E_s$ (for a single bridge), there is obtained

$$\frac{E_d}{E_{do}}=\left\{\sqrt{2}\left[\frac{(3+\sqrt{3})(2X_c^2-\sqrt{3}X_s^2)-(3-\sqrt{3})X_cX_s}{4X_c^2-3X_s^2}\right]\sin\frac{\delta}{2}+(2+\sqrt{3})\sin\left(15°-\frac{\delta}{2}\right)\right\}\times\cos\left(\alpha_m+\frac{\delta}{2}+15°\right) \tag{29}$$

which can be written by substituting $u=\delta+30°$, as shown on Fig. 8

$$\frac{E_d}{E_{do}}=\frac{1-\sqrt{3}}{2\sqrt{2}}\left[\frac{2+k\sqrt{3}}{2-k\sqrt{3}}\sin\left(\frac{u}{2}-15°\right)-(2+\sqrt{3})\cos\left(\frac{u}{2}-15°\right)\right]\times\cos\left(\alpha_m+\frac{u}{2}\right) \tag*{(30)*}$$

Fourth Mode of Operation

The period of simultaneous commutations between the two bridge circuits cannot exceed 60 degrees, since at $\delta=30$ degrees the commutating angle is 60 degrees, as shown in Fig. 9. When the commutating angle of a single bridge circuit reaches 60 degrees, further increase in the reactance factor will introduce forced delay within the bridge circuit.[1] Similarly, in the case of 12-phase operation, an increase in the reactance factor beyond the value for which $\delta=30$ degrees will introduce an additional forced delay of angle β within each bridge. Thus, as shown in Fig. 10, there will be a total delay of $\alpha_m+\beta$ in the start of commutation, followed by a commutation angle of 60 degrees. In this mode of operation, the maximum value of β is 30 degrees. The limiting value of $\alpha_m+\beta$ ($\beta\leq 30°$) is 60 degrees, at which point the direct voltage is zero.

There are two commutating intervals involved in this mode of operation: in the first interval, $\omega t=\alpha_m+\beta$ to $\omega t=\alpha_m+\beta+30°$, commutations from T_9 to T_1 and T_8 to T_{12} occur simultaneously; in the second interval, $\omega t=\alpha_m+\beta+30°$ to $\omega t=\alpha_m+\beta+60°$, commutations from T_9 to T_1 and T_{10} to T_2 occur simultaneously.

Currents in the first interval are determined by equations 16 and 17, in the second interval by equations 19 and 20. Performing the integrations the same way as in deriving equation 23

$$\frac{I_dX_c}{E_s}=\frac{\sqrt{6}X_c}{2X_c+\sqrt{3}X_s}\sin(\alpha_m+\beta+30°)$$

which can be written

$$\frac{I_dX_c}{E_s}=\frac{\sqrt{6}}{2+k\sqrt{3}}\sin(\alpha_m+\beta+30°) \tag*{(31)*}$$

The instantaneous voltages in the two intervals are given by equations 25 and 27, respectively. Following the same procedure used in deriving equation 30

$$\frac{E_d}{E_{do}}=\sqrt{3}\times\left[\frac{(2X_c^2-\sqrt{3}X_s^2)-(2-\sqrt{3})X_cX_s}{4X_c^2-3X_s^2}\right]\times\cos(\alpha_m+\beta+30°) \tag{32}$$

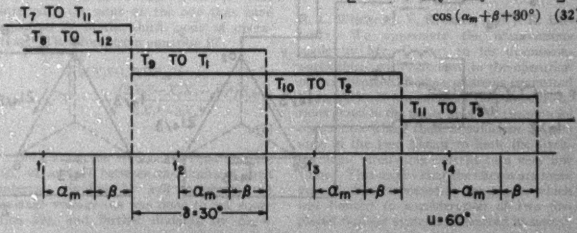


Fig. 10. Graph showing interval of commutations during the fourth mode of operation

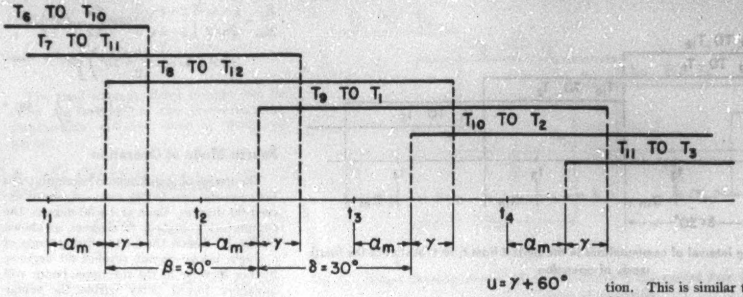


Fig. 11. Graph showing interval of commutations during the fifth mode of operation

which can be written

$$\frac{E_d}{E_{do}}=\frac{\sqrt{3}(1-k)}{2-k\sqrt{3}}\cos(\alpha_m+\beta+30°) \qquad (33)^*$$

Fifth Mode of Operation

Equations 31 and 33 apply for values of reactance factors and ratio of source to commutating reactance for which $\alpha_m+\beta\leq 60°$ and $\beta\leq 30°$. Thus, for $\alpha_m\leq 30°$, β increases over the full range of 0 to 30 degrees, and the fourth mode of operation terminates at $\beta=30$ degrees. When $\alpha_m=30$ degrees, the voltage ratio is zero when $\beta=30°$. This represents the dividing point, as for $\alpha_m>30°$, the voltage ratio becomes zero before $\beta=30°$.

For ratios of source reactance to commutating reactance less than 2/3, there is a fifth mode of operation. This mode is analogous to the third mode of operation of a 6-phase double-way[1] (single bridge), in that there are intervals when three phases commutate simultaneously, resulting in a temporary 3-phase short-circuit.

During the period of commutation from T_9 to T_1, there are five distinct intervals, as shown in Fig. 11. In the first interval $\omega t=\alpha_m+\beta$ to $\omega t=\alpha_m+\beta+\gamma$, commutations exist from T_7 to T_{11} and T_9 to T_1 (in bridge I) and from T_8 to T_{12} (in bridge II). In the second interval, $\omega t=\alpha_m+\beta+\gamma$ to $\omega t=\alpha_m+\beta+30°$, commutations exist from T_9 to T_1 (in bridge I) and T_8 to T_{12} (in bridge II). During the third interval, $\omega t=\alpha_m+\beta+30°$ to $\omega t=\alpha_m+\beta+30°+\gamma$, there are commutations from T_9 to T_1 (in bridge I) and T_8 to T_{12} and T_{10} to T_2 (in bridge II). In the fourth interval, $\omega t=\alpha_m+\beta+30°+\gamma$ to $\omega t=\alpha_m+\beta+60°$, there are two commutations, T_9 to T_1 (in bridge I) and T_{10} to T_2 (in bridge II). During the fifth interval, $\omega t=\alpha_m+\beta+60°$ to $\omega t=\alpha_m+\beta+60°+\gamma$, there are three commutations, T_9 to T_1 and T_{11} to T_3 (in bridge I) and T_{10} to T_2 (in bridge II). In the fifth mode, the angle $\beta=30°$.

In Fig. 12, the current i_1 is the instantaneous current corresponding to i_{1c}, the commutating current for a T_9 to T_1 commutation. This is similar to current i_1 of Fig. 5. Also shown in the figure are the commutating currents i_3 and i_{3c}, corresponding to a T_7 to T_{11} commutation, and i_4 and i_{4c}, corresponding to a T_8 to T_{12} commutation. Not shown in the figure are similar currents for other commutations involved: i_2 which flows in terminal c and out terminal a during a T_{10} to T_2 commutation; and i_5 which flows in terminal c and out terminals a and b during a T_{11} to T_3 commutation. These currents will be used in the derivations which follow:

FIRST INTERVAL: $\omega t=\alpha_m+\beta$ to $\omega t=\alpha_m+\beta+\gamma$.

Commutations are T_7 to T_{11} and T_9 to T_1 (bridge I) and from T_8 to T_{12} (bridge II). Therefore

$$E_{ab}=-\frac{3}{2}Lpi_1-\frac{3}{2}Lpi_3-2L_spi_4+E_{a1b1}$$

$$=-\frac{3}{2}L_spi_1-\frac{3}{2}L_spi_3-2Lpi_4+E_{a2b2}$$

$$E_{bc}=\frac{3}{2}Lpi_3+L_spi_4+E_{b1c1} \qquad (34)$$

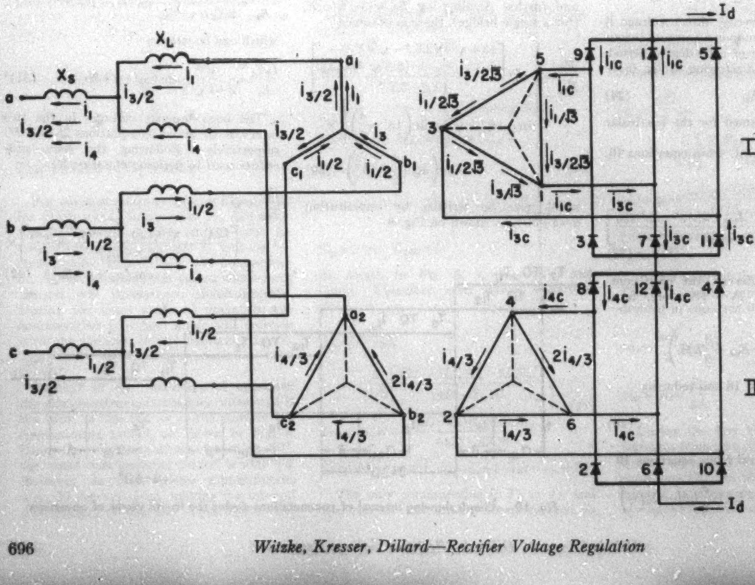


Fig. 12. Circuit diagram showing currents during the first interval of the fifth mode of operation

$$E_{a1b1}=E_{a2b2}=E_{b1c1}=0$$

As shown in Fig. 12, $i_{1c}=(\sqrt{3}/2)i_1$, $i_{3c}=(\sqrt{3}/2)i_3$, $i_{4c}=i_4$. Therefore, after solving equations 34

$$pi_{1c}=\frac{\sqrt{3}}{2}\,pi_1=-\frac{1}{\sqrt{3}}\times\left[\frac{(2L+L_s)E_{ab}+2(L+L_s)E_{bc}}{2L(L+L_s)}\right] \quad (35)$$

SECOND INTERVAL: $\omega t=\alpha_m+\beta+\gamma$ to $\omega t=\alpha_m+\beta+30°$.

The only commutations are T_9 to T_1 (bridge I) and T_8 to T_{12} (bridge II). Therefore, from equations 17

$$pi_{1c}=\frac{\sqrt{3}}{2}\,pi_1=-\frac{1}{\sqrt{3}}\times\left[\frac{(2L-3L_s)E_{ab}-2LE_{ca}}{(4L^2-3L_s^2)}\right] \quad (36)$$

THIRD INTERVAL: $\omega t=\alpha_m+\beta+30°$ to $\omega t=\alpha_m+\beta+30°+\gamma$.

There are commutations from T_9 to T_1 (bridge I) and T_8 to T_{12} and T_{10} to T_2 (bridge II). Therefore

$$E_{bc}=-L_spi_2+L_spi_4+E_{b1c1}$$
$$=-Lpi_2+Lpi_4+E_{b2c2}$$

$$E_{ca}=\frac{3}{2}Lpi_1+2L_spi_2+L_spi_4+E_{c1a1} \quad (37)$$

$$E_{a1}=0,\ E_{a1b1}=E_{c1a1}=-\frac{E_{b1c1}}{2},\ E_{b2c2}=0$$

After solving equations 37

$$pi_{1c}=\frac{\sqrt{3}}{2}\,pi_1=\frac{1}{\sqrt{3}}\left[\frac{E_{bc}+2E_{ca}}{2(L+L_s)}\right] \quad (38)$$

FOURTH INTERVAL: $\omega t=\alpha_m+\beta+30°+\gamma$ to $\omega t=\alpha_m+\beta+60°$.

There are commutations from T_9 to T_1 (bridge I) and T_{10} to T_2 (bridge II). Therefore, from equations 20

$$pi_{1c}=\frac{\sqrt{3}}{2}\,pi_1=-\frac{1}{\sqrt{3}}\times\left[\frac{2LE_{ab}-(2L-3L_s)E_{ca}}{4L^2-3L_s^2}\right] \quad (39)$$

FIFTH INTERVAL: $\omega t=\alpha_m+\beta+60°$ to $\omega t=\alpha_m+\beta+60°+\gamma$.

There are three commutations, T_9 to T_1 and T_{11} to T_3 (bridge I) and T_{10} to T_2 (bridge II). Therefore

$$E_{ab}=-\frac{3}{2}Lpi_1-L_spi_2+E_{a1b1}$$

$$E_{bc}=-L_spi_2-\frac{3}{2}Lpi_3+E_{b1c1}$$

$$E_{ca}=\frac{3}{2}Lpi_1+2L_spi_2+\frac{3}{2}Lpi_3+E_{c1a1} \quad (40)$$
$$=\frac{3}{2}L_spi_1+2Lpi_2+\frac{3}{2}L_spi_3+E_{c2a2}$$

$$E_{a1b1}=E_{b1c1}=E_{c1a1}=E_{c2a2}=0$$

After solving equations 40

$$pi_{1c}=\frac{\sqrt{3}}{2}\,pi_1=\frac{1}{\sqrt{3}}\times\left[\frac{(2L+L_s)E_{ca}+2(L+L_s)E_{bc}}{2L(L+L_s)}\right] \quad (41)$$

Following the same procedure used in obtaining equation 23 during the third mode of operation and substituting $\beta=30°$, $\gamma=u-60°$, $k=X_s/X_c$

$$\frac{I_dX_c}{E_s}=\frac{\sqrt{2}}{(1+k)(2+k\sqrt{3})}\left[\sin\left(\frac{u}{2}-30°\right)+\sqrt{3}(1+k)\cos\left(\frac{u}{2}-30°\right)\right]\times\sin\left(\frac{u}{2}+\alpha_m+30°\right) \quad (42)^*$$

The average direct voltage is obtained, following the same procedure used in obtaining equation 30, by averaging the instantaneous voltages over a 60-degree period of symmetry. It is necessary to consider only the first four current intervals since they extend over a 60-degree period, and a second symmetrical voltage pattern starts at the beginning of the fifth interval. Taking bridge I, and referring to Fig. 6(B)

$$e_d=-E_b+E_1=\sqrt{3}E_{b1} \quad (43)$$

where E_{b1} is determined for the particular interval considered.

In the first interval, from equations 34, during the T_7 to T_{11} commutation

$$e_{d1}=\sqrt{3}E_{b1}=0 \quad (44)$$

In the second interval, the only commutations are T_9 to T_1 and T_8 to T_{12}; therefore, equation 16 applies. Thus

$$e_{d2}=-\sqrt{3}\left[\frac{(2X_c^2-X_cX_s-X_s^2)E_{ab}+2(X_c^2-X_s^2)E_{ca}}{4X_c^2-3X_s^2}\right] \quad (45)$$

This is the same as equation 25.

In the third interval, commutations exist from T_9 to T_1, T_8 to T_{12}, and T_{10} to T_2. Therefore, from equations 37

$$e_{d3}=\sqrt{3}E_{b1}=\frac{\sqrt{3}}{2}\left[E_{bc}+L_spi_2-L_spi_4\right]$$

which can be solved with equations 37 to yield

$$e_{d3}=\sqrt{3}\left[\frac{(L-L_s)E_{bc}}{2L}\right] \quad (46)$$

In the fourth interval, the only commutations are T_9 to T_1 and T_{10} to T_2. This interval is the same as the one that gave equation 27 in the third mode of operation. Therefore, from equation 27

$$e_{d4}=-\sqrt{3}\left[\frac{2(X_c^2-X_s^2)E_{ab}+(2X_c^2-X_cX_s-X_s^2)E_{ca}}{4X_c^2-3X_s^2}\right] \quad (47)$$

Integrating the voltages of equations 44, 45, 46, and 47 between the limits of each interval, dividing by $\pi/3$ to obtain the average voltage (as was done to get equation 29), and further dividing by $E_{do}=(3\sqrt{6}/\pi)E_s$, there is obtained after substituting $\beta=30°$, $\gamma=u-60°$, and $k=X_s/X_c$

$$\frac{E_d}{E_{do}}=\frac{\sqrt{3}(1-k)}{2-k\sqrt{3}}\left[\cos\left(\frac{u}{2}-30°\right)-\sqrt{3}(1+k)\sin\left(\frac{u}{2}-30°\right)\right]\times\cos\left(\frac{u}{2}+\alpha_m+30°\right) \quad (48)^*$$

References

1. INFLUENCE OF A-C REACTANCE ON VOLTAGE REGULATION OF 6-PHASE RECTIFIERS, R. L. Witzke, J. V. Kresser, J. K. Dillard. *AIEE Transactions*, vol. 72, pt. I, July 1953, pp. 244–53.

2. EXTENDED REGULATION CURVES FOR 6-PHASE DOUBLE-WAY AND DOUBLE-WYE RECTIFIERS, I. K. Dortort. *AIEE Transactions*, vol. 72, pt. I, May 1953, pp. 192–202.

3. THREE-PHASE RECTIFIER CIRCUITS, A. J. Maslin. *Electronics*, New York, N. Y., vol. 9, Dec. 1936, pp. 28–31.

4. MERCURY-ARC RECTIFIERS AND THEIR CIRCUITS (book), D. C. Prince, F. B. Vogdes. McGraw-Hill Book Company, Inc., New York, N. Y., 1927.

5. *Discussion* by R. L. Witzke, J. V. Kresser, and J. K. Dillard of reference 2, pp. 198–202.

◆

Discussion

I. K. Dortort (I-T-E Circuit Breaker Company, Philadelphia, Pa.): It is indicated in the paper that at $K=0$ (no common reactance) the 12-phase circuit shown in Fig. 5 behaves in the same manner as the 6-phase double-way circuit. This is confirmed by the limiting curves of Figs. 1, 2, and 3. For $K=0$ and highly inductive loads the curves are exactly the same as obtained from two earlier articles (refs. 1 and 2 of the paper). With no common a-c reactance the two 6-phase systems must be entirely independent, and can have no effect on each other.

When two displaced 6-phase systems are connected in parallel, the inductance of the d-c system is common to both and will maintain the total current at constant value without preventing the flow of sixth-harmonic current between the two. The derivations of this paper will apply to two displaced 6-phase double-way rectifiers in parallel, if each is separately equipped with sufficient d-c inductance, or an interphase transformer is placed between them to hold the sixth-harmonic circulating current to a very small value.

R. L. Witzke, J. V. Kresser, and **J. K. Dillard:** We appreciate the observations made by Mr. Dortort in his discussion, particularly with reference to the operation of two displaced 6-phase systems connected in parallel. It is significant that the derivations given in the paper apply to the parallel connection where there is sufficient inductance in the load circuit to limit the sixth-harmonic circulating current to a very low value. This means that the curves are more general than indicated in the paper, which was limited to consideration of two displaced 6-phase systems connected in series.

Heat Transfer from Electron Tubes at High Altitudes and High Ambient Temperatures

MRS. B. O. BUCKLAND
MEMBER AIEE

THIS paper contains the results of a series of temperature tests conducted on a *6AQ5* beam power tetrode at full rating at simulated altitudes and ambient temperatures. The effects of altitude, shielding, tube position, and ambient temperature on temperature rise of the hottest spot of the glass envelope are shown. A few measurements were made of the effect of surrounding a single tube by a cluster of other tubes of the same type and rating. The tube was not delivering any output.

The temperature at which a tube or unit may operate must be limited if adequate life is to be attained. High plane speeds and altitudes increase operating temperatures by imposing a higher ambient temperature and by reducing the convective heat transfer. Also, tubes must often be shielded to protect them from vibration and shock, further reducing the heat transfer and increasing the severity of operating conditions. Evaluation of these altitude, ambient temperature, and shielding effects is necessary for the design of air-borne equipment.

Effect of Altitude, Excluding Higher Ambient Temperature

Test results at room temperature and altitudes up to 123,000 feet for several types of shielding and mounting plates, in both vertical and horizontal positions, are summarized in Fig. 1. Ambient temperature was measured by a thermocouple inside the bell jar.

SHIELDING

Fig. 1 shows that:

1. Blackened shields result in much lower hot-spot temperature of the glass envelope than shiny shields.
2. Use of a slotted shield, instead of a full shield, further reduces temperature of the tube envelope.
3. The unshielded tube runs the coolest, although not much below the slotted shield. See also Fig. 2.

TUBE POSITION AND MOUNTING

1. Mounting the tube horizontally results in a hot spot at the top of the tube. This hot spot is higher than the corresponding hot spot with vertical mounting.
2. Tube temperatures appear to be somewhat lower if the tube is mounted on a blackened metal plate, but the effect of the mounting plate on temperature is small compared to the effect of shielding; see Fig. 1.

Effect of Altitude, Including High Ambient Temperature

An increased ambient temperature reduces the envelope temperature rise but always increases the actual temperature. This effect is indicated in Figs. 3, 4, and 5. The average of three thermocouples embedded in the oven walls was used as the external ambient temperature.

Effect of Enclosure

Close proximity of the tube to the enclosure walls results in an appreciably higher temperature rise of the tube envelope. This is shown in Fig. 4, where test points for the lowest ambient temperature (room temperature) were obtained without any oven around the tube, whereas all tests at higher ambient temperatures were made with the small oven enclosing the tube. Fig. 8 shows that in the latter case, the distance from the envelope of a centrally located tube to the oven wall is about 1.7 inches. In Fig. 4 the curves through the higher ambient test points are extrapolated to room temperature to permit comparison with test points taken without the oven.

In actual practice, tubes are likely to be installed even closer to the walls of the enclosure. Thus, the results here reported from tests in the oven are as significant as, or even more significant than, the tests without the oven. For closer proximity of the tube to an enclosing wall the temperature rises for the various conditions here considered may be higher than the present curves would indicate.

Proximity of Other Tubes

Limited data on the operation of tubes in a cluster indicate an appreciable tem-

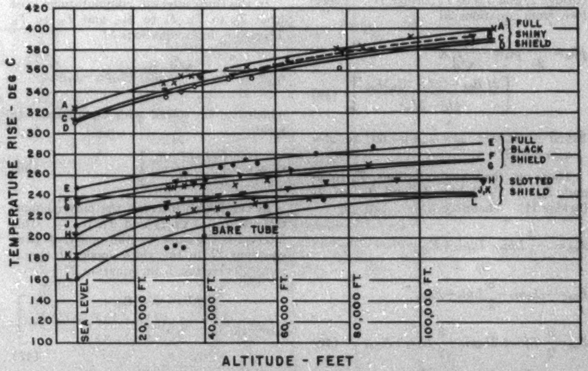


Fig. 1. Effect of shielding on temperature rise of glass envelope hot spots at various altitudes. Data taken at room ambient. Tests made in vacuo; pressure converted to altitude from Fig. 11

A, G, K, L—Tube vertical on blackened metal plate
B, D, F, J—Tube horizontal on blackened metal plate
C—Tube vertical on shiny metal plate
E, H—Tube vertical on textolite plate

Paper 53-307, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 20, 1953; made available for printing May 4, 1953.

MRS. B. O. BUCKLAND is with the General Electric Company, Schenectady, N. Y.

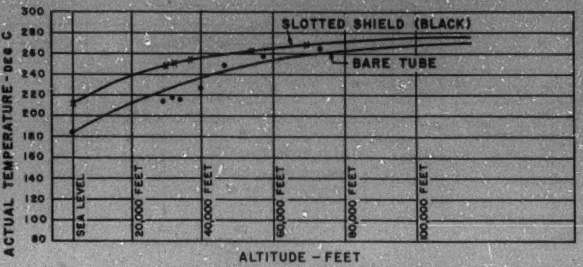


Fig. 2. Comparison of glass envelope hot spots unshielded and with slotted black shield at altitude. Tube vertical on blackened metal plate

perature rise over that found in a single tube operation; see Figs. 6 and 7. Therefore, in using the quantitative data in this report, care should be taken to allow for any proximity effects that may exist due to the presence of other heat sources.

Test Procedure

Measurement of envelope hot-spot temperatures was made for the *6AQ5* at full rating under a variety of conditions. Means had to be provided for measuring temperature accurately, holding vacuum, holding input power constant and measuring it, and attaining high ambient temperatures. Since the tube output was not under investigation, the anode and screen were tied together, with a separate circuit for the heater. The tube rating is 16.85 watts, made up of 14.00 watts for anode and screen together, and 2.85 heater watts. All data were corrected to this rating before being plotted.

All tests were carried out in a 12-inch-diameter bell jar held on a steel plate by vacuum sealing grease. The steel plate was perforated in several places to accommodate various connections: vacuum pump and manometer connections, tube power supply, thermocouple connections, and oven heater connections. All plugs for electrical connections were vacuum tight, being the standard feed-through terminals for hermetic sealing.

For higher ambient temperatures, an electrically heated oven was used inside the bell jar. Fig. 8 shows the tube cluster arrangement inside the oven. For single-tube tests, the central position was used.

MEASUREMENT AND CONTROL

Temperature Measurement

Due to the high temperatures expected, chromel-copnic thermocouples were used. Their outputs were read with a Leeds and Northrup potentiometer. The wires for measuring the glass envelope temperature were 5 mils in diameter; all others were 15 mils. The 5-mil size was used because preliminary tests showed that conduction along the wires caused the 15-mil couples to read some 35 to 40 degrees centigrade (C) less than the 5-mil size for a rise of the order of 150 C. The cold junction was immersed in an ice and water bath.

Experimentation with various techniques showed that the best way of fastening the couple to the glass consisted of bending the wires adjacent to the bead into an arc. Then the thermocouple was tied to the envelope with glass thread at two positions about 1/2 inch apart. The springiness of the bent wire is sufficient to hold the bead firmly against the glass. Wire of this fine thickness does not appreciably disturb the free-convection boundary layer.

Room ambient temperatures were measured by a thermocouple inside the bell jar, located some distance away from the tube under test. Higher ambient temperatures were measured by three thermocouples inside the oven. The oven thermocouples were fastened to the oven walls by peening them into the material on the inside. For the single-tube vertical tests, the thermocouple was placed opposite the anode center. For the single-tube horizontal tests, the thermocouple was placed at the top of the tube, above the anode center.

Vacuum Control and Measurement

The vacuum system of Fig. 9 was used in all cases. The vacuum was set by the valve in the line between pump and bell jar, with the pump operating just enough to take care of the system leakage after the desired vacuum was reached. Pressure was measured with a Bennert manometer. The maximum vacuum that the system would hold with the valve wide open was 0.1 centimeter of mercury.

Input Power Control and Measurement

The tube power supply was controlled by a voltage regulator. Current and voltage to the plate and screen were measured with *DP-2* meters while heater current and voltage were measured with a *P-3* ammeter and a Simpson *V.O.M.*

High Ambient Temperature Control

The electric furnace, Fig. 8, was used to simulate high ambient temperatures. This furnace, which contained the tube or tubes, was placed in the bell jar. The several heating elements were wound for the same watts per square inch power density, and the wall temperatures measured. Adjustments to secure uniform temperature were made by placing variable resistors in the circuits and adjusting these until the three thermocouples read very nearly the same. The oven wall thermocouple temperature averages were used as the external ambient temperature.

PRELIMINARY TESTS

The position of the hot spot was explored by fastening couples at intervals

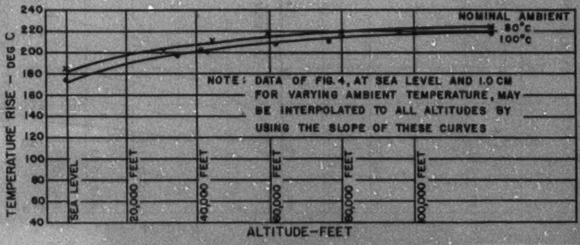


Fig. 3. Effect of higher ambient temperature on glass envelope hot spot at altitude. Tube with slotted black shield, mounted on blackened metal plate

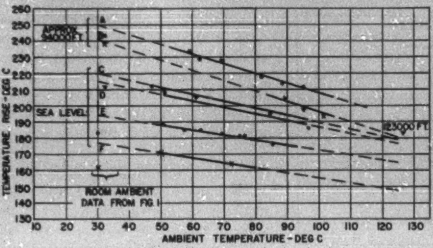


Fig. 4. Variation in temperature rise of glass envelope hot spot with ambient temperature change

A, E—Tube vertical, slotted shield
B, F—Tube vertical, bare
C—Tube horizontal, slotted shield
D—Tube horizontal, bare

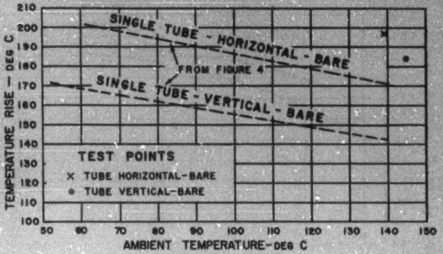


Fig. 6. Effect of the proximity of other heat sources at sea level, for varying ambient temperatures. Tube in center of cluster of seven tubes

along the length of the tube. The location was found to be at approximately the center of the anode, as shown in Fig. 10.

Heating curves were run to determine the time to reach substantial equilibrium. The results were:

Bare tube, no oven, 8 minutes
Shielded tube, no oven, 15 minutes
Oven tests, 2 hours

CORRECTION TO RATED POWER DISSIPATION

Throughout the experimental procedure the total input power was to be held at the rated value of 16.85 watts. However, small deviations occurred in most of the tests. A correction was applied so that all results would represent behavior at rated input.

Where the deviation from rated power was small, less than 4 per cent, the corrected temperature rise was taken as proportional to the power input, without introducing an appreciable error.

For the few cases of power deviation greater than 4 per cent, account had to be taken of the variation in heat-transfer coefficients with temperature. Since in all these cases the power input was low, the corrections were less than would be given by a linear power-ratio multiplication. Test points as plotted on all curves have been corrected to 16.85 watts input.

Analysis of Data and Results

The temperature data are plotted against altitude rather than air pressure in centimeters, although pressure is the parameter regulated in the laboratory. The altitudes were converted from data[1] of Navord Report *1488*, plotted on Fig. 11.

In all cases, an increase in altitude increases the temperature rise and actual temperature of the tube envelope. This is primarily due to the reduction in heat transfer by free convection as altitude is increased; see Fig. 10.

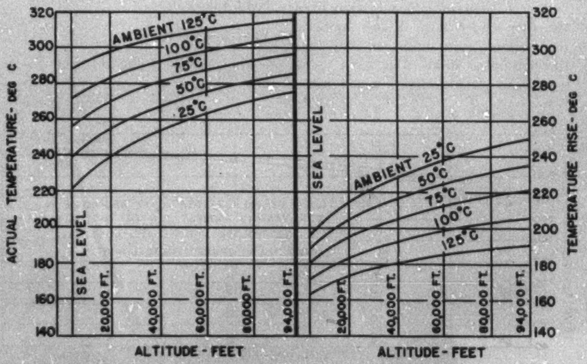


Fig. 5. Illustrative curves showing the effect of ambient temperature on both temperature rise and actual temperature, at altitude, tube vertical, slotted shield

ROOM-AMBIENT SINGLE-TUBE DATA

The variation in temperature rise with shielding, altitude, and tube position is qualitatively in accord with theory. Heat transfer by free convection should vary directly as the square root of the air density and by radiation directly as the emissivity. Departures from this simple theory are accounted for by the introduction of an air resistance between bulb and shield, and by the high conductivity of the shield material, which distributes the concentrated input power at the glass envelope over a wider area of the shield. The slotted black shield causes temperature rises that are substantially lower than for any other shield because most of the envelope surface remains uncovered. The full shiny shield causes a greater temperature rise than the full black shield because radiation from the shiny shield is much less.

Inspection of Figs. 1 and 2 shows that addition of a shield to the bare tube has more effect on the temperature rise at low altitudes than at high altitudes.

Heat is transmitted from the tube to the shield in two parallel paths; by radiation and by conduction through a narrow air space. From the shield it flows to the surroundings by convection and radiation, again in parallel. At sea level heat flow by convection is much greater than it is at high altitudes. Therefore the total resistance to flow at high altitudes is higher than it is at sea level, with a consequent reduction in the percentage of the resistance of the total circuit due to the shield. Thus the percentage increase in temperature due to the shield becomes less as the altitude increases.

The data show that hot spot temperatures are higher if the tube is mounted horizontally rather than vertically and that the difference decreases as altitude

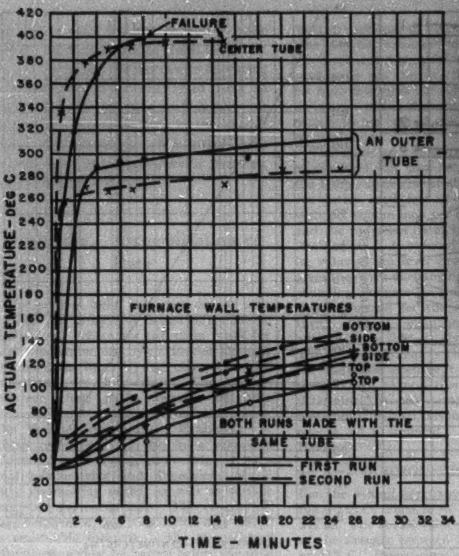


Fig. 7 (left). Test at altitude of tubes in cluster. Center tube operated to failure. Tubes vertical, bare. Air pressure 0.2 centimeter mercury. Altitude approximately 123,000 feet

Fig. 8 (above). Electrically heated oven used to simulate high ambient temperature thermocouples peened into inside walls. One at top, one at side wall, one at bottom

increases. For the case of uniform circumferential heat generation, a tube mounted vertically has a hot spot of constant temperature around the circumference of the tube. In the horizontal position a hot spot, shown by the data to be of higher temperature, exists at the top surface. Since the difference in hot spot temperatures is due to a difference in free convection, it decreases with increasing altitude.

Higher-Ambient Single-Tube Data

The curves of Figs. 3 and 4 show the trend of decreasing temperature rise with increasing ambient temperature. This is expected theoretically since both radiation and free convection heat transfer increase with increasing ambient temperature.

Because two hours were required to reach equilibrium for each point, only two tests at higher ambients were run over the complete range of altitudes. Attempts at using shorter intervals resulted in curves whose slopes were the same as those of Fig. 3 but whose temperature rises did not follow the expected trend with increasing ambient temperature. The points of Fig. 4 were taken at 2-hour intervals, but at sea level and at one high vacuum point only. The data for any ambient temperature may be interpolated to include all altitudes by using the slopes of the curves on Fig. 3.

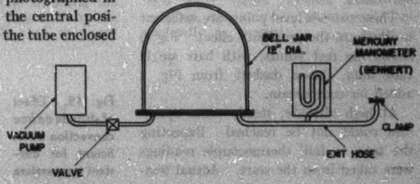
While temperature rise decreases with increasing ambient temperature, actual temperature increases. The difference between sea level and high altitude rises, Fig. 4, decreases as ambient temperature is raised. Moreover, the effects of tube position and of shielding tend to become less as the ambient increases. All these results are due to increasing heat transfer by radiation with increasing ambients, and to decreasing heat transfer by convection with increasing altitude.

All the test points made at room ambient, both for sea level and at 1.0 centimeter air pressure, from Fig. 1, are included in Fig. 4 as room temperature data. These points lie well below the straight lines of test results for higher ambient extrapolated down to room ambient.

This anomaly illustrates an extremely important concept, namely, the effect on the temperature of close proximity of the tube to the enclosure wall. All the room ambient runs were made in a large bell jar, while the higher ambient runs were made in the small oven photographed in Fig. 8 with the tube in the central position. At sea level, with the tube enclosed in the oven, and with no oven power, the tube heated the walls to 50 C. Therefore, no room ambient point could be taken with the tube enclosed in the oven. It it had been possible to cool the walls to room ambient, or 30 C, the temperature rise would have followed the straight line. This means that enclosing the tube causes an additional rise at sea level of 16 C for the bare tube above the original 160 C rise and 11 C for the slotted shield. If it had been possible to heat the bell jar walls to the desired ambients, the rise of the tube would parallel the present straight lines, would lie considerably below them, and would include the room ambient points. This rise would represent the effect of a single tube in a larger space at the specified ambient. In cases where a single tube is thus located, its temperature rise will be substantially less than the curves of Fig. 4, but in most cases the tube will be enclosed, and the curves as drawn are more realistic.

With the information from Figs. 3 and 4, two families of illustrative curves are plotted for the vertical tube with slotted shield (Fig. 5). The family on the left

Fig. 9. Schematic diagram of the vacuum system used to simulate high altitude pressure



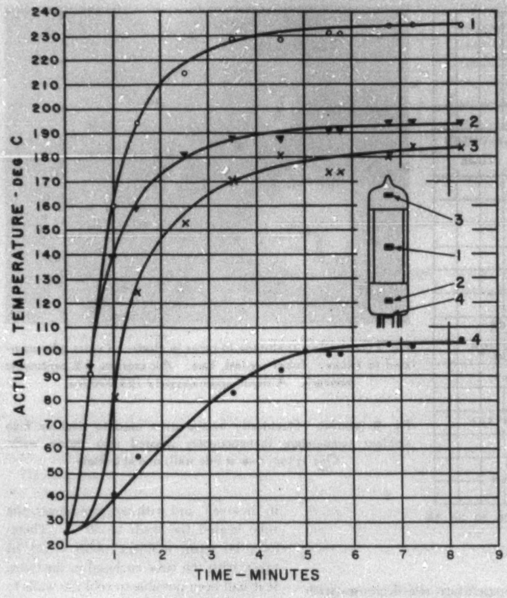


Fig. 10. Temperature data locating the tube envelope hot spot

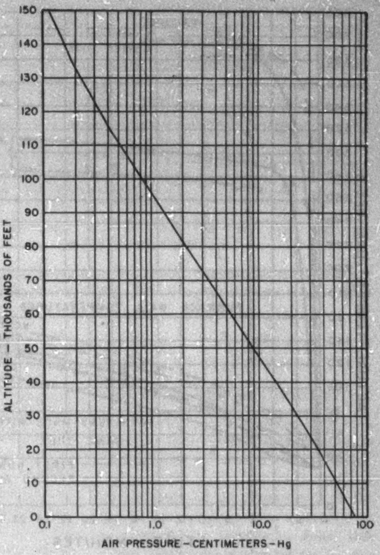


Fig. 11. Relation between altitude and pressure[2]

show temperature rise, and it will be noted that the rise is highest for the lowest ambient. The family on the right show actual temperature, and here the total temperature is highest for the highest ambient. Both families show that the temperature rises as the altitude increases.

Tubes in Cluster

Operation of the tube under test in the center of a circle of six similar tubes shows a substantially greater temperature rise in cluster than in single-tube operation. This was to be expected since the presence of the surrounding tubes increases the temperature of the air and of the enclosing surface. At sea level the seven tubes heated the walls to about 140 C with no power on the furnace. Consequently only two points could be taken, one with the tube vertical and the other with it horizontal.

These two sea level points are sufficient to illustrate the proximity effect. Fig. 6 shows the test points, with bare single tube data, shown dashed, from Fig. 4 added for comparison.

At high altitude, steady state operation could not be reached. Expecting the tube to fail, thermocouple readings were taken from the start. Actual temperatures as functions of time are plotted in Fig. 7. From the observed difference between center tube and outer tube, it appears that proximity effects are even greater at altitude than at sea level.

It is interesting to note that both runs in Fig. 7 were made with the same tube. After the first run, when failure occurred, the apparatus was allowed to cool, and a second run was taken, as the tube appeared to recover. Again failure occurred. In both cases, the tube failed at approximately 400 C. It no longer functioned after the second run.

Included in Fig. 7 is a temperature curve for one of the tubes in the surrounding circle. Its rise is comparable with that of a single tube at a corresponding ambient. The thermocouple was fastened on the outer side, nearest the oven wall, and thus presumably did not measure the hottest spot. Apparently the hot spot will exist on the side adjacent to the center tube. Because of the poor conductivity of the glass, a fairly large circumferential gradient is likely to exist between the point of measurement and the actual hot spot.

The limited information on tubes in

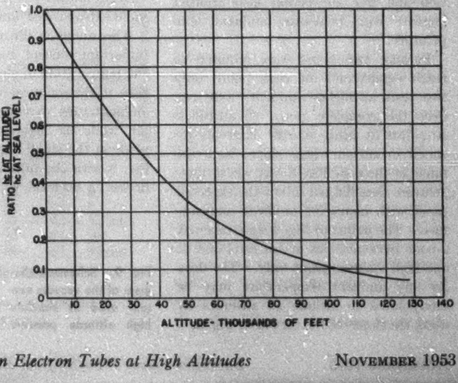


Fig. 12. Effect of altitude on free convection coefficient for constant temperature

clusters is valid for the particular tube spacing and enclosure of these tests. It clearly illustrates the increase in operating temperature which results from using an enclosed bank of tubes. However, to generalize from these data in the case of other configurations, with other heat sources such as transformers or resistors, might lead to serious error.

Appendix. Generalization of Data

To predict bulb hot spot for a tube during the process of design offers severe difficulties. There are two main paths for heat dissipation, one by conduction through the base, the other by free convection and radiation from the surface. The division between these paths depends on tube geometry and materials, which determine the thermal resistances. Evaluation of these resistances for use in a complete thermal circuit requires knowledge of thermal conductivity and dimensions of all heat conducting paths, as well as surface heat-transfer coefficients. The circuit will be complicated, but its solution may afford a reasonable estimate if the temperature distribution must be known before the tube is built.

Even the fairly complete information of the present tests offers difficulty when an analytical study of the bulb temperature is attempted. Using the rated input of 16.85 watts and glass bulb area of 4.75 square inches, the power density for the bare tube is 3.55 watts per square inch. Referring to the well-known heat-transfer relations,[2] the bulb temperature is calculated to be much higher than the test figure. Therefore, a substantial portion of the heat must be dissipated elsewhere. Assuming that both the measured temperature rise at sea level and the published heat-transfer coefficients are correct, the power density for heat flow from the surface is 2.45 watts per square inch. This means that approximately 30 per cent of the heat is dissipated by conduction.

This figure of 2.45 watts per square inch at sea level can now be used to calculate the temperature rise at altitude. The free-convection coefficient at sea level is 0.008 watt per square inch per degree centigrade, and the radiation coefficient is 0.0072. At 60,000 feet, h_c is reduced to 27 per cent of its value at sea level (from Fig. 11, which gives the effect of changing pressure without changing temperature) or 0.0022. A further correction for the effect of temperature, without changing power input, amounting to 0.95, is given in reference 2. The final coefficient is then 0.0021. Since the radiation coefficient is a function of temperature, several tries are necessary to determine the new temperature. As the first approximation, assume the rise is 225 C over a 25 C ambient. Then $h_r=0.0098$. Hence

$$t_w-t_a=2.45/(0.0021+0.0098)=206\text{ C}$$

For the second approximation, assume the rise is 210 C. Then, $h_r=0.0092$, and $t_w-t_a=$ 217 C. The actual value may then be taken as the average between 217 C and 206 C, or 212 C. Comparison with Fig. 1 shows that the test temperature rise at 60,000 feet is 222 C, exceeding the calculated temperature by 10 C.

A similar calculation at 120,000 feet gives 232 C instead of the test of 240 C. The same procedure can be carried out for higher ambient temperatures, where the principal influence is on the radiation coefficient. This procedure should then be reasonably satisfactory for determining the effect of altitude and ambient temperature on any other bare tube, when the temperature at sea level is known.

When it comes to predicting the effect of shielding on the glass envelope temperature, the method is more cumbersome. The previously outlined procedure will give the temperature of the shield provided the emissivity is known, but not the temperature of the glass bulb enclosed within it. The bulb is separated from the shield by a dead air space, so that heat is transmitted by conduction through the air as well as by radiation to the shield. To evaluate these resistances in the thermal circuit between bulb and shield requires air gap dimensions and air conductivity, as well as shield emissivity. For the shiny shield, the principal path of heat flow is by conduction through the air, resulting in a substantially lower bulb temperature than will be the case without this conduction. At high altitudes, fortunately, the conductivity of the air remains constant, being practically independent of pressure in this pressure range. Thus, while free convection outside the shield falls off rapidly with increasing altitude, conduction between bulb and shield does not; in fact, the conductivity of air increases with temperature. A rough calculation was made of bulb temperatures, with 1/16-inch air gap, thermal conductivity 0.0009 watts per inch per degree centigrade, for shield emissivities of 0.9, for the black shield, and 0.4 for the shiny shield. The result for the black shield at sea level was 272 C and for the shiny shield 358 C. Comparing these results with the tests of Fig. 1 shows that the method gives conservative results, as both calculations exceed the tests by some 30 C.

References

1. HANDBOOK OF SUPERSONIC AERODYNAMICS. *Navord Report No. 1488*, Washington, D. C., 1950.

2. BASIC HEAT-TRANSFER DATA IN ELECTRON TUBE OPERATION, Mrs. B. O. Buckland. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1079–85.

◆

Discussion

E. R. G. Eckert (University of Minnesota, Institute of Technology, Minneapolis, Minn.): The author has covered the field which she set out to investigate very thoroughly and in the Appendix is given a theoretical treatment which the complex nature of the setup permits.

Conditions with higher ambient air temperature are investigated by placing an oven around the arrangement of electron tubes. The question arose in my mind in this connection whether the situation shown in Fig. 4 that an extrapolation of the higher temperature curves down to room temperature does not agree with direct measurements at room temperature is due only to the proximity effect which the author mentions or partially also due to the situation that the oven wall temperatures may be higher than the temperature of the air between the electron tubes and the oven walls.

In the Appendix, the author mentions that a heat transfer between the electron tube and its shield would occur by conduction through the air. I wonder whether the air space between tube and shield is closed off sufficiently to prevent a convection flow caused by the fact that the air in this space is probably warmer than the air outside the shield (chimney effect).

Mrs. B. O. Buckland: Prof. Eckert's remarks about the effect of enclosing the tube in a small oven are very much to the point. It is not possible that the oven wall temperatures were higher than the temperature of the air in the oven, as the oven was completely enclosed with heating elements on top and bottom as well as on the sides. During the test, a thermocouple was placed in the air between the tube and the wall, and in all cases it measured some temperature intermediate between tube and walls. Because its temperature was a function of its position in the boundary layer, its use as a reference temperature was discarded.

The air space between the envelope and shield of a *6AQ5* is approximately 1/16 inch. At the base there is no room for air flow, so for this particular case the chimney effect is absent. However, for larger tubes with wider spaces between shield and envelope, the chimney effect would undoubtedly be present.

The Effect of Minor Constituents in High Dielectric Constant Titanate Capacitors

W. W. COFFEEN
NONMEMBER AIEE

Synopsis: The potentials and limitations of high dielectric constant capacitors are briefly outlined. Additives used to modify the dielectric properties of $BaTiO_3$ are listed and classified into two groups: 1. the "shifters," whose additions primarily cause a shift in the temperature of the Curie peak, and 2. the "depressors," whose additions result primarily in a depression of the Curie peak. The effect of $BaSnO_3$ and $PbSnO_3$ additions are used to illustrate the behavior of the shifter group, and the variations in the performance of the other members of this group are discussed. Similarly, the effect of additions of $MgSnO_3$, $Bi_2(SnO_3)_3$, $MgTiO_3$, and $CaTiO_3$ are shown as typical of the behavior of members of the second group. Generalizations are offered regarding the effect of composition on nonlinearity and aging, and these are illustrated.

BARIUM titanate ceramic capacitors were revealed in the postwar period as the newest gem of the capacitor industry. As in most gems, however, when the eyes were no longer dazzled by the glitter, in this case, of its unusually high dielectric constant, some flaws began to appear. During the years since the war, ceramic capacitor manufacturers have been busily engaged in learning to improve and augment the sterling qualities of barium titanate by the use of additives while their partners in circuitry, the electrical engineers, have sought to find new and different applications to take advantage of the unique properties of these high dielectric constant ceramic materials.

The data of this paper are presented in an effort to acquaint the electrical engineer with the tools of the trade of the ceramic engineer in formulating barium titanate capacitors. It is hoped that a clearer understanding of both the potentials and limitations of barium titanate capacitors will result, giving rise to a more rapid and more satisfactory utilization of these materials in electric circuits.

Present Limitations of Titanate High Dielectric Constant Capacitors

The range of application of high dielectric constant capacitors in the electronic industry is indicated in Fig. 1, representing data taken from P. S. Schmidt.[1] High dielectric constant capacitors are restricted commercially to a rather narrow but very important range of capacitance from about 100 to 20,000 micromicrofarads. Crystalline ceramic dielectrics as a whole are limited in their use in capacitors by two important physical considerations. First, no satisfactory commercial process has yet been devised for inexpensive production of multiple-layer capacitors. Second, for cheapest production processes, thicknesses are limited to about 20 mils. Thicknesses down to about 5 mils may be produced by somewhat more costly methods[2,3] but weaknesses due to the increased voltage gradients tend to restrict the use of the high dielectric constant bodies in the thinner films.

The ranges of dielectric constant required for various sizes of capacitors using plates 0.005 inch and 0.020 inch thick with commercially acceptable areas are shown in Fig. 2. As Fig. 2 implies, bodies may be compounded having dielectric constants as high as 10,000. However, the higher the dielectric constant, the more restricted is the operating range. For example, *K*-10,000 bodies will have minimum dielectric constants of 10,000 over a temperature range of about 10 to 15 degrees centigrade; *K*-6,000 bodies may have dielectric constants of 6,000 or above over a range of about 30 to 40 degrees centigrade; *K*-3,000 bodies may retain this value as a minimum over a range of 60 to 70 degrees centigrade, while *K*-1,000 bodies may be obtained which will be operable over a range of 140 degrees centigrade from −55 to 85 degrees centigrade. When additional requirements on the variation of the dielectric constant above the minimum value must be met, the operating ranges may be reduced still further. The result is that for capacitors of 20-mil thickness, 20,000 micromicrofarads is about the upper limit of capacitance which can be obtained for capacitors capable of any general application. Thinner capacitors might be used, of course, to obtain capacitance values well above 20,000 micromicrofarads but, unfortunately, as will be noted later, the bodies with higher dielectric constants are much more sensitive to variation from bias voltages, and this effect is magnified to such an extent by the increased voltage gradient in the thin films that only very special use can be made of the bodies of highest dielectric constant in the thin films.

From this discussion, it is obvious that two general methods of approach to the improvement of high dielectric constant capacitor bodies are open: 1. to develop bodies of at least an intermediate level of dielectric constant, but with improved performance under the higher voltage gradients, or 2. to develop bodies of higher dielectric constant which exhibit less variation in dielectric constant with temperature. Both approaches are being utilized commercially today.

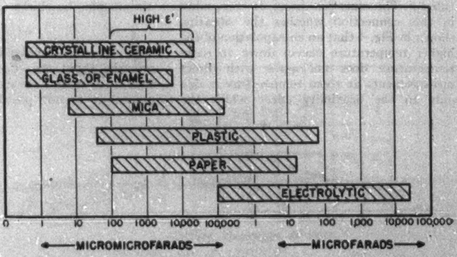


Fig. 1. Capacitance ranges for various types of capacitors

Paper 53-236, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Summer General Meeting, Atlantic City, N. J., June 15–19, 1953. Manuscript submitted March 17, 1953; made available for printing April 30, 1953.

W. W. Coffeen is with the Metal and Thermit Corporation, Rahway, N. J.

The author wishes to express his appreciation to G. L. Dillman, W. J. Suto, and Fred Does, Jr., who prepared and measured the stannate-titanate specimens; to Dr. L. R. Bickford, Alfred University, Dr. R. R. Shively, Jr. Zenith Radio Corporation, and J. L. Jenkins, Radiation, Inc., for permission to utilize data from personal communications; and to J. W. Deaderick, American Lava Corporation, for his critical review.

Table I. Classification of Additives

Shifters	Depressors
$BaSnO_3$	$MgSnO_3$
$SrSnO_3$	$NiSnO_3$
$CaSnO_3$	$Bi_2(SnO_3)_3$
$PbSnO_3$	$MgZrO_3$*
$CuSnO_3$	Fluorides of Ba, Ca, Sr, Mg, etc.
$ZnSnO_3$	$CaTiO_3$
$CdSnO_3$	$MgTiO_3$
$BaZrO_3$	
$CaZrO_3$	
$SrZrO_3$	
$SrTiO_3$	
$PbTiO_3$	

* $MgZrO_3$ is generally recognized to be a solid solution of MgO and ZrO_2 in a 1-to-1 mole ratio, rather than a compound.

Before discussing the use of additives to $BaTiO_3$, it is well to review the characteristics of $BaTiO_3$ itself. Fig. 3 shows the variation of dielectric constant and dissipation factor as a function of temperature according to von Hippel.[4] Examination of these data shows that $BaTiO_3$ has a dielectric constant of about 1,400 and dissipation factor of about 0.0150 at room temperature. These characteristics are reasonably constant in the range of 25 to 85 degrees centigrade and $BaTiO_3$ would be satisfactory in this respect for some applications. However, for use at temperatures above 85 or below 25 degrees centigrade, the variations in dielectric constant become very large; at the lower temperatures, high dissipation factors are encountered. However, the short service life of capacitor bodies containing only $BaTiO_3$ is perhaps their most serious disadvantage. While dielectric strengths of 100 volts per mil or more have been reported for $BaTiO_3$ disks at room temperature, failure frequently occurs in a few hours at a temperature of 85 degrees centigrade with voltage gradients as low as 50 volts per mil.

Insulation resistance, initially of the order of 10^{13} megohms, decreases very rapidly under such life test conditions. Fortunately, service life is improved substantially by almost any of the additives used to produce changes in dielectric constant and dissipation factor, so that the latter criteria become dominant in the initial stage of development of an improved body. One other characteristic of $BaTiO_3$, which is also common to many of the high dielectric constant titanate bodies, is the variation in their properties as a function of voltage gradient. In general, the dielectric constant is increased slightly as the initial voltage is applied, and then decreased rapidly as the voltage gradient is increased. The initial increase is normally more pronounced with alternating than with direct

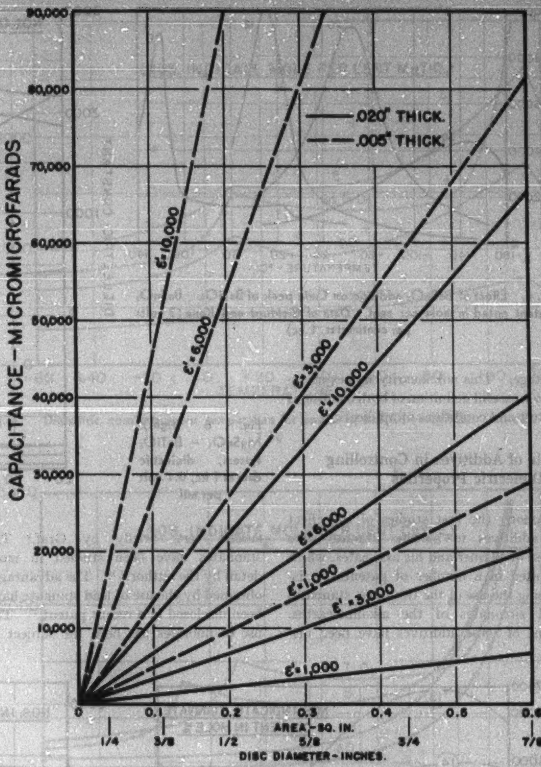


Fig. 2. Capacity versus area for ceramic disks of various ε' values

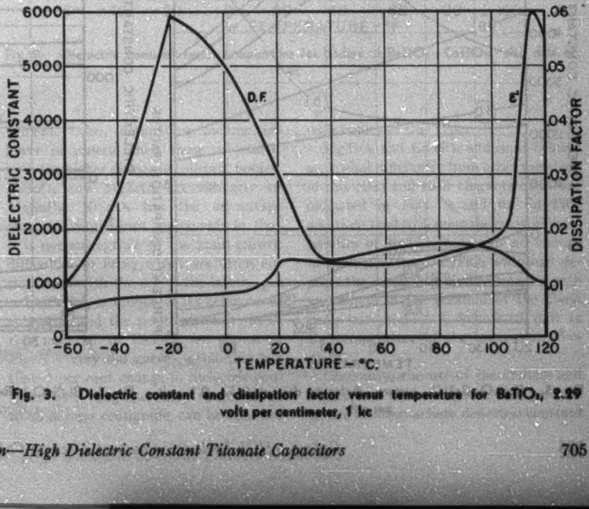


Fig. 3. Dielectric constant and dissipation factor versus temperature for $BaTiO_3$, 2.29 volts per centimeter, 1 kc

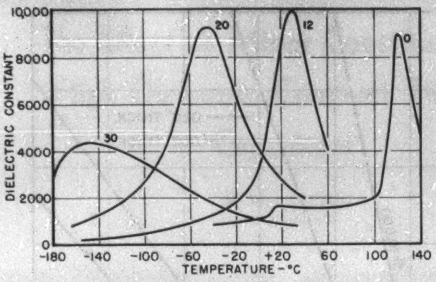


Fig. 4. Effect of $BaSnO_3$ addition on Curie peak of $BaTiO_3$. $BaSnO_3$ content noted in mole per cent. Data of Bickford and Kane (3 volts per centimeter, 1 kc)

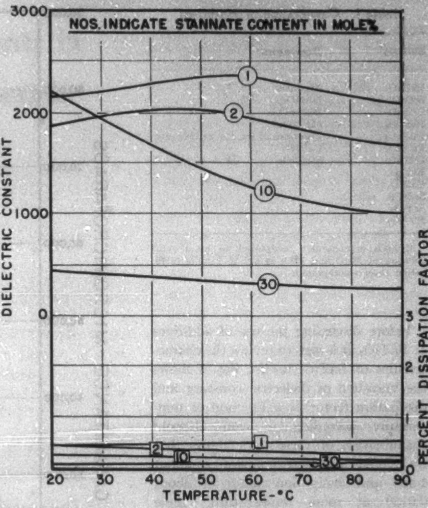


Fig. 6 (right). $MgSnO_3 - BaTiO_3$ system, dielectric data at 1 kc, 0.1 volt per mil

voltage. This nonlinearity makes necessary a careful definition of both conditions of test and conditions of application.

Role of Additives in Controlling Dielectric Properties

Among the first studies of the effect of additives to barium titanate were those of Wainer and his associates, which resulted in a number of patents[5-7] disclosing the use of the titanates, stannates and zirconates of the alkaline-earths. Some of these additives have been discussed very briefly by Graf.[8] The stannates have been studied in more detail by the author.[9, 10] The advantages obtained by the use of lead stannate have been disclosed in a recent patent.[11] The use of fluorides has been the subject of two British patents.[12, 13] Simple oxides such as MgO, BaO, TiO_2, MnO_2, Fe_2O_3, ZrO_2, SnO_2, and others have been used in very small quantities, primarily to produce minor changes or to improve firing behavior.

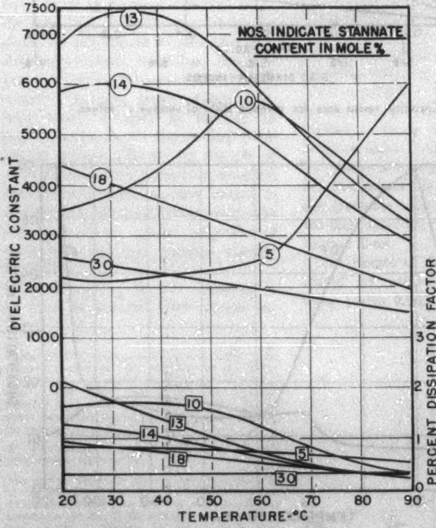


Fig. 5. $PbSnO_3$–$BaTiO_3$ system, dielectric data at 1 kc, 0.1 volt per mil

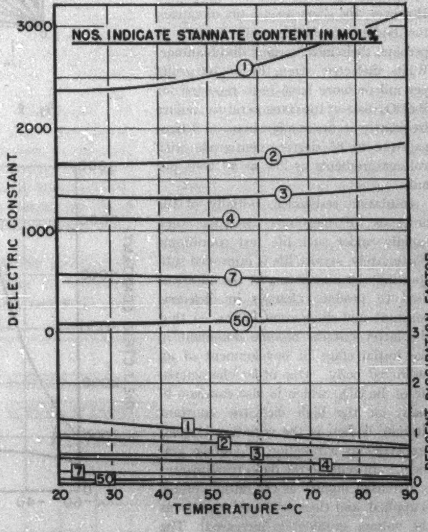


Fig. 7. $Bi_2(SnO_3)_3$–$BaTiO_3$ system, dielectric data at 1 kc, 0.1 volt per mil

The additives can be classified into two general types. One type which we can call the shifters has the effect of shifting the temperature of the Curie peak of barium titanate, normally downward. The second type can be called depressors and, in general, this type tends to depress the Curie peak, although small additions of some of the members of this class increase the dielectric constant at temperatures below the original Curie temperature, while depressing the value of ϵ' at temperatures in the vicinity of 120 degrees centigrade. There are variations within each group, as we shall see, but they are divided roughly as shown in Table I.

The behavior of a shifter type of additive is illustrated by the behavior of $BaSnO_3$, shown in Fig. 4. These curves are based on unpublished data presented by Bickford and Kane at the 15th Symposium on Ceramic Dielectrics held in April 1952 at Rutgers University. Reference to the figure shows the continuous drop in Curie temperature as the amount of $BaSnO_3$ in the body is increased. The downward shift is accompanied by a broadening of the peak and a change in the peak dielectric constant, this value increasing with small additions of $BaSnO_3$, then decreasing again.

The other shifters, with the exception of $PbTiO_3$, behave in a similar manner, but differ in degree. $SrTiO_3$ shifts the peak most slowly per mole added and has the sharpest peaks. Stannates in general shift the peak most rapidly per mole added. Zirconates are intermediate in rate of shift. Peak values of dielectric constant are highest for barium, calcium, and strontium stannate additions, and these additives are virtually interchangeable on a molar basis.[10] $PbSnO_3$ bodies exhibit the flattest curves of any of the shifters. Fig. 5 shows the results of composition changes in the system: $BaTiO_3$–$PbSnO_3$. $PbTiO_3$ is unique in that addition of this material produces an increase in the Curie temperature with accompanying decrease in dielectric constant values in the lower temperatures. Bodies containing $CuSnO_3$ are characterized by comparatively high dissipation factors and $CdSnO_3$ bodies are difficult to fire, so neither of these has any commercial interest.

The behavior of the depressor type of additive is illustrated in Fig. 6, showing the changes made by varying composition in the $BaTiO_3$–$MgSnO_3$ system. The levels of dielectric constant are considerably lower (<2,500) here than those obtained by the use of shifters (up to 10,000), but much improved stability is evident. Also, dissipation factors are lower in general than those associated with the higher dielectric constant bodies. $NiSnO_3$ and $MgZrO_3$ act similarly to $MgSnO_3$. $NiSnO_3$ has the advantage over the magnesium compounds in that it is not susceptible to the grain-growth difficulties in firing which are often encountered with the use of the latter.

$Bi_2(SnO_3)_3$ is unique in its behavior in this class and the result of additions of this material to $BaTiO_3$ are shown in Fig. 7. Very flat curves, exhibiting less than 1 per cent change in dielectric constant over the operating range of 20 to 85 degrees centigrade, can be obtained with bodies of this type.

$MgTiO_3$ and $CaTiO_3$ additions behave somewhat differently from other members of this class and their characteristics are indicated in Figs. 8 and 9. $MgTiO_3$ additions tend to depress the curve characteristics of $BaTiO_3$ but with no change in peak position. $CaTiO_3$ additions depress the curve in a similar manner with no change in the position of the Curie peak but produce a downward shift in the temperature of the lower transition point.

Obviously, the use of the shifters and depressors in combination to produce bodies of intermediate dielectric constant

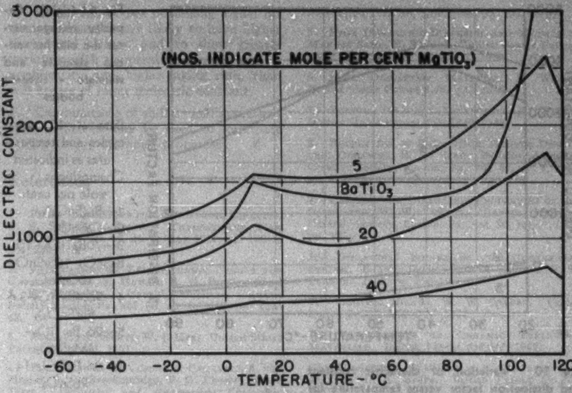


Fig. 8. Dielectric constant versus temperature for bodies of $BaTiO_3 + MgTiO_3$. All data at 1 kc

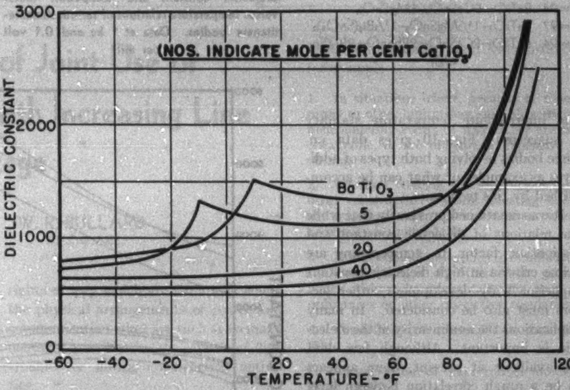


Fig. 9. Dielectric constant versus temperature for bodies of $BaTiO_3 + CaTiO_3$. All data at 1 kc

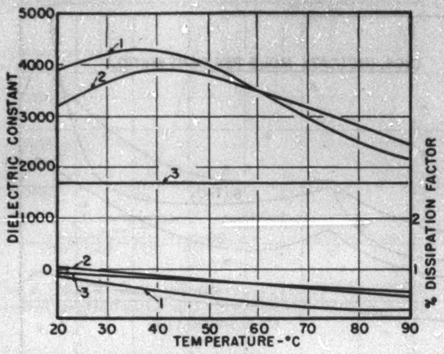


Fig. 10. Variation of dielectric constant and dissipation factor versus temperature for mixed stannate-$BaTiO_3$ bodies at 1 kc, 0.1 volt per mil

Compositions in mole per cent:

1 = 94 $BaTiO_3$–5$CaSnO_3$–1$MgSnO_3$
2 = 97 $BaTiO_3$–1$^1/_2$$MgSnO_3$–1$^1/_2$$Bi_2(SnO_3)_3$
3 = 94 $BaTiO_3$–2$CaSnO_3$–2$PbSnO_3$–2$Bi_2(SnO_3)_3$

Fig. 11 (right). Capacity change versus d-c bias for various stannate and stannate - titanate bodies

Data at 10 megacycles and temperatures as indicated Compositions in mole per cent:

1—$Bi_2(SnO_3)_3$
2—$CaSnO_3$
3—70$BaTiO_3$–10 $BaSnO_3$–10$SrSnO_3$–10$CaSnO_3$
4—96.4$BaTiO_3$-3.6 $Bi_2(SnO_3)_3$
5—96 $BaTiO_3$–4$MgSnO_3$
6—86 $BaTiO_3$–14 $PbSnO_3$

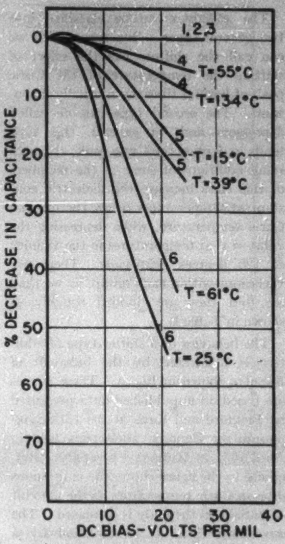


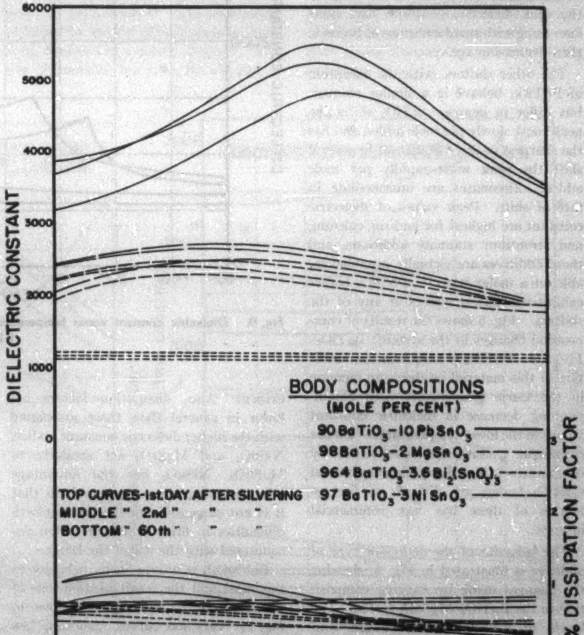
Fig. 12 (below). Aging characteristics of dielectric constant and dissipation factor versus temperature relations of various stannate-titanate bodies. Data at 1 kc and 0.1 volt per mil

and intermediate temperature stability is suggested. Fig. 10 gives data on three bodies involving both types of additives as examples of what can be accomplished by this technique.

It was mentioned previously that while the relations of dielectric constant and dissipation factor to temperature are prime criteria in high dielectric constant capacitor body development, other factors must also be considered. In many applications the nonlinearity of the dielectric is important. Although few data are available at present, there appears to be a rough correlation between level of dielectric constant and degree of nonlinearity as shown in Fig. 11. In the group of bodies shown, the $PbSnO_3$ body has the highest dielectric constant, the $MgSnO_3$ body next, followed in turn by the $Bi_2(SnO_3)_3$ body, the alkaline-earth stannate body, and finally $Bi_2(SnO_3)_3$ and $CaSnO_3$ themselves.

Aging rates likewise seem to increase in general with an increase in level of dielectric constant as shown in Fig. 12.

Although the potential properties of a high dielectric constant capacitor are determined by a proper choice of additives, it should be recognized that these properties are often not realized in production without the aid of very careful adjustment and control of processing. The various steps in manufacture such as mixing, forming, firing, silvering, attach-



ment of leads, and final coating all require study to insure that the finished capacitor will have the properties indicated by laboratory tests. Performance of dielectric bodies in production trials is often equal in importance to the dielectric properties themselves in dictating a satisfactory choice of additives.

Conclusions

1. The minor constituents of high dielectric constant titanate capacitor bodies play a major role in determining the dielectric properties of the bodies.
2. All of the dielectric properties are susceptible to alteration, but not independently.
3. Additives to $BaTiO_3$ bodies may be divided into shifters, which produce primarily bodies having very large dielectric constants, and depressors, which produce primarily bodies with lower dielectric constants.
4. In general, bodies possessing higher dielectric constants are likely to have higher dissipation factors, and to show greater variation with temperature change, greater nonlinearity and higher aging rate, than will bodies of lower dielectric constant.
5. Combinations of shifter- and depressor-type additives may be used to formulate bodies having optimum properties.

References

1. Fixed Capacitors for Electronic Circuits, P. S. Schmidt. *Electrical Manufacturing*, New York, N. Y., vol. 47, no. 5, May 1951, pp. 100–5, 244, 246, 248.

2. Fabrication of Thin Ceramic Sheets for Capacitors, G. N. Howatt, R. G. Breckenridge, J. M. Brownlow. *Journal*, American Ceramic Society, Easton, Pa., vol. 30, no. 8, 1947, pp. 237–42.

3. Dielectric Sheets, W. J. Lies. United States Patent *2,539,446*, 1951.

4. High Dielectric Constant Ceramics, A. von Hippel, R. G. Breckenridge, F. G. Chesley, Laslo Tisza. *Industrial and Engineering Chemistry*, Easton, Pa., vol. 38, Nov. 1946, pp. 1097–1109.

5. High Dielectric Material and Method of Making Same, E. Wainer. United States Patents *2,402,515*, *2,402,516*, *2,402,517*, 1946.

6. High Dielectric Material, E. Wainer. United States Patent *2,452,532*, 1948.

7. Barium Titanate Dielectric, E. Wainer. United States Patent *2,436,839*, 1948.

8. Properties of Some High Titania Dielectric Ceramics, E. G. Graf. *Bulletin*, American Ceramic Society, Easton, Pa., vol. 31, no. 8, 1952, pp. 279–82.

9. Ceramic and Dielectric Properties of the Stannates, W. W. Coffeen. *Journal*, American Ceramic Society, Easton, Pa., vol. 36, July 1, 1953, pp. 207–14.

10. Dielectric Bodies in the Quaternary System: $BaTiO_3$–$BaSnO_3$–$SrSnO_3$–$CaSnO_3$, W. W. Coffeen. *Ibid.*, pp. 215–19.

11. Ceramic Bodies, W. W. Coffeen. United States Patent *2,624,709*, 1953.

12. Improvements in Ceramic Dielectric Materials. British Patent *669,941*, 1952.

13. Dielectric Materials and Method of Making, C. E. Every. British Patent *586,802*, 1947.

No Discussion

Some Aspects of Joint Use of Wood Pole Lines with Increasing Line Voltage

W. E. BLOECKER
MEMBER AIEE

W. R. BULLARD
FELLOW AIEE

THE joint use of wood poles by supply and communication companies has been an established way of doing business for well over 50 years. Because of changing technologies in both the power and telephone fields the successful co-ordination of facilities on the same pole line requires continuous review. It is the purpose of this paper to discuss the problems which were introduced by the trends to higher supply voltages.

Before going into that part of the subject, it will be well to review some of the background. The joint use of wood poles or, as it is more commonly referred to, joint use, is the use of the same pole line to support the wires or cables of both the power companies and the communication companies. A special term, joint crossing pole, is applied to the case where the wires of the two utilities are attached to the same pole at a point where their routes intersect. As the discussion progresses, it will be recognized that many of the considerations of joint use apply equally well to what are called conflicting rights of way which are situations where the physical arrangements of power and communication wires are such as to make contacts possible in the event of line troubles or during construction activities.

It is clear that the joint occupancy of poles, or situations involving numerous crossings of power and communication circuits, require added engineering consideration to assure the safety of the public, of utility workmen and of property, comparable to that afforded by separate nonconflicting plants. What then are the reasons for entering into joint use construction?

There are several important advantages which a single-pole line offers over separate lines under many circumstances for example:

1. In situations where, because of topography or other limitations, suitable separate nonconflicting rights of way either are not obtainable, or are difficult or expensive to obtain.
2. Accurate control of separations and other construction features affecting safety at crossings and other locations where proximity is unavoidable.
3. Elimination of objections to double lines by municipalities or other civic groups which ultimately might take the form of ordinances or other restrictive regulations.
4. Possibilities of economies in construction.
5. Conservation of pole timber now in short supply.

For reasons such as these, the joint use of pole lines to carry supply and communication circuits, started before 1900, has developed to the point where there are in the United States some 11,000,000 poles doing this double duty, not counting the cross poles. Of these jointly used poles, there are about 8,500,000 in urban areas, and about 2,500,000 in rural areas. Since 1945 the total number of jointly used poles has been increasing at the rate of about 500,000 each year.

In the beginning, practically all of the situations for which joint use appeared justified involved power voltages of less than 5 kv, and the protection arrangements of the telephone plant were designed accordingly. As a result, 5 kv came to be considered the maximum sup-

---

Paper 53-317, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Pacific General Meeting, Vancouver, B. C., Canada, September 1–4, 1953. Manuscript submitted February 4, 1953; made available for printing July 1, 1953.

W. E. Bloecker is with the American Telephone and Telegraph Company, New York, N. Y.; W. R. Bullard is with Ebasco Services, Inc., New York, N. Y.

ply voltage with which joint construction could safely be established. The greater part of urban joint construction today involves voltages less than 5 kv. However, with increasing loads, the power companies found it necessary to employ higher voltages. The continuation and extension of joint construction with these higher voltages involved a reappraisal of the reaction on noise induction and protection. That these considerations have been effectively worked out may be seen from the fact that most of the rural joint use today involves distribution voltages of 12 kv, and extending as high as 25 kv, while urban joint use with 13.2 kv is fairly common in some areas. There are a growing number of situations of joint use with 33 kv, some scattered cases of joint use with 44 kv (5 miles in one case), with 69 kv (2 miles in one case), and there is one situation where arrangements were worked out for something over 400 feet of telephone cable to be joint with 110 kv.

This is not to say that all caution has been cast aside but rather that, where proper co-ordinated protection has been applied, both to the power and to the communication systems, joint use may well represent the best engineering solution without reference to any arbitrary voltage limitation.

Noise Considerations

As far as inductive co-ordination is concerned, it might appear that the higher couplings resulting from the relatively close spacing of power and communication conductors add to the difficulties. However, the fixed relationship between the two types of plant results in a more uniform exposure which facilitates control of the coupling. In general, noise problems are no more difficult to solve than where the power and communication lines are at roadway separation.

In considering the effect of increasing voltage on power system influence, that is, its potentiality for impairing transmission on neighboring communication facilities, it may be said that under some conditions it will increase, and under others it will decrease.

The circumstances under which power circuit influence is likely to go up when voltages increase are:

1. When there are relatively large harmonics of the higher orders impressed on the distribution system usually at the source of power supply.
2. Where the loads are relatively light, at least during certain periods of the day.

The circumstances under which noise influence is likely to go down are:

1. Where the wave shape of the power supply system is fairly good and the harmonic currents are of the lower orders: 3, 5, 7, and 9.
2. Where the transformer kilovolt-amperes per mile of line is fairly high, the reason being that transformer magnetizing currents are likely to control current wave shape on lower voltage lines, and these currents decrease as the voltage increases.

In any event, the correction of an inductive situation involving a high-voltage distribution circuit is generally less difficult than one involving a low-voltage circuit for two reasons:

1. Harmonic currents tend to be smaller on the higher impedance circuits.
2. Where harmonic currents are large enough to cause trouble, it is much easier to reduce their effect because shunt devices for high impedance circuits are easier to design and cheaper to build than similar devices for low-impedance circuits.

Protection Considerations

The approach to successful, safe operation of power and telephone circuits on the same poles or involving numerous crossings rests on the following basic considerations:

1. Adequate structural arrangements to minimize the possibility of contacts occurring, and to provide safe working conditions.
2. Co-ordinated electrical protection arrangements for power and telephone facilities to minimize hazards to persons, and to limit plant damage in the event of accidental contact.

The broad basis for structural arrangements rests on the generally accepted standards of good engineering as set forth in the National Electrical Safety Code published by the National Bureau of Standards and approved by the American Standards Association. It is the result of contributions by engineers of the wire-using companies, and reflects their experience and best judgment at any given state of electrical development. This code, however, is no immutable law. In fact, it is now in its fifth edition, as industry has learned from experience what changes are desirable or necessary, and has recognized the latest technical developments. Organizational work is now under way looking to a new edition, which will include consideration of those aspects relating to voltage classifications and clearances, strengths of structures, and items which have called for extensive interpretation since the fifth edition was published.

The code allows considerable latitude in interpretation, and to arrive at sound construction practices the Joint Committee on Plant Co-ordination, sponsored by the Edison Electric Institute and the Bell Telephone System, has issued specification *M-12*[9] which, in general, outlines methods for attaching the two plants to the poles, and gives the requirements for pole strengths, guying, wire sags and tensions, and clearances, both above the ground and between the two plants. Specifications which provide for adequate working and climbing space for the workmen of both utilities are also included. As an example of keeping up with a changing art, proposals for further savings in pole timber by judicious reduction of clearance requirements have been made, and it is the thought that, as experience is gained with these proposals, they could ultimately be incorporated into *M-12* with such changes as seem desirable. Because of its nature, *M-12* has been incorporated in many joint use contracts, frequently by reference, in order to permit changes resulting from technical developments to be incorporated in *M-12* and consequently in the construction practices, without reopening the contract itself.

Adequate electrical protection involves, basically:

1. Prompt and positive de-energization of the power conductors in the event of contact with communication plant.
2. Establishment and maintenance of good low-resistance grounds on the communication plant.

Prompt de-energization is important to limit the duration of hazardous voltages on the communication plant so as to minimize the possibility of personal injury, and to limit the extent of plant damage. The low-resistance ground is important to assist in obtaining prompt de-energization, and also to minimize the magnitude of voltage rise on the communication plant before de-energization. Thus, it will be seen that the electrical protection involves the practices of both the power and communication systems, and the co-operative engineering and application of proper protection devices to the two systems is what is meant by the term co-ordinated protection.

Co-ordinated Electrical Protection

Co-ordinated electrical protection involves consideration of a number of factors, some of which relate to the power system, and others to the communication plant, so that each particular situation needs to be reviewed as an individual case reflecting the particular conditions. These factors are discussed in the following.

Voltage of Power System

From the standpoint of prompt and positive de-energization it can be stated as a general principle that the higher the supply voltage, the better the situation. A couple of homely arithmetical examples will illustrate this point:

1. Assume a 4-kv circuit (2.4 kv to neutral) with 250-kva load per phase. The power load current would be slightly over 100 amperes, which would probably mean that fuses or circuit breakers set to operate at about 200 amperes would be provided. In the event of accidental contact with telephone plant such that the total impedance of the fault circuit is around 10 ohms, the fault current would be about 240 amperes, giving a small margin of current to operate fuses or circuit breakers.

2. With an increase to 12.5 kv (7.2 kv to neutral) with a load of 750 kva per phase, the full-load current again would be of the order of 100 amperes. However, an accidental contact would result in a fault current of 550 amperes (making allowance for the somewhat higher reactance of the supply transformer) thereby providing a substantial margin to insure prompt de-energization.

Type of Power System

The type of power system distribution has an important bearing on the assurance that de-energization will be prompt and positive under fault conditions. The multigrounded neutral system is the most favorable from this standpoint, since the frequent grounding of the neutral conductor, together with the neutral conductor itself, affords a low impedance path for fault currents in the event of faults to neutral, or to other metallic structures or circuits associated with the neutral. The multigrounded neutral also provides by suitable bonding, a ready means of obtaining a good ground for telephone cables and protectors.

In case of a unigrounded system, whether the neutral conductor is carried along with the phase wires or not, a fault to ground completes a circuit back to the substation through the earth. The impedance of this fault current path may be relatively high, depending to a great extent on the ground contact resistance.

In the case of isolated delta systems in the event of a fault to ground of a single-phase conductor, the capacitance to ground of the unfaulted phase conductors provide the only path for fault currents. Therefore, in a single-phase fault to ground, the fault current is the leakage and charging current of the unfaulted conductors. Generally, the smaller the system and the lower the voltage, the smaller will be the resulting fault current. The questions to be answered in this case are whether or not such fault current is sufficient to operate the power system protection equipment, and if not, whether the voltage on the communication plant is held low enough through low impedance grounds to be safe.

The occurrence of a single fault on an isolated delta system increases the voltage to ground of the nonfaulted phase wires, and the breakdown of another conductor often follows. In the event of a double fault, the situation is similar to that of the unigrounded system. In some delta systems grounding banks are used at the supply substation. From the standpoint of co-ordinated protection, this arrangement modifies the isolated delta system to a form more or less like that of the unigrounded system, depending upon the characteristics of the grounding means.

Type of Telephone Plant

Two types of telephone plant, namely, cable and open-wire, may be involved in joint use with power systems.

In the case of cable, the impedance of any fault to ground is kept low by frequent bonding of the sheath to the neutral conductor of a multigrounded neutral system where such a system is present. In other cases, a ground may be obtained by the connection of the cable sheath to the sheath of buried or underground cables. Bonding to other extensive underground structures, such as water pipe systems, is also often effective in lowering the impedance to ground. With either of these methods of grounding to extensive underground systems, the cable sheath corrosion aspects need to be given careful consideration.

For open-wire situations where voltages above 3 kv to ground are encountered, carbon gap devices designed to break down at about 3 kv are connected between each wire and ground at approximately 1/2-mile intervals. A feature of the design of these devices is their ability to handle large amounts of current. For example, in the case where the power system reclosers operate three times before de-energizing permanently, the device is able to carry three successive shots of 600 amperes of 1-second duration each, and up to 1,300 amperes for durations of 2/10 second. To be effective, the resistance of the ground connections must be low and here again the multigrounded neutral, where available, is most useful.

Characteristics of Power Protective Systems

Power protective devices may have several forms and may, particularly in the case of transmission circuits, involve rather complex arrangements. For distribution circuits, however, they usually consist of fuses or circuit breakers, the operation of which depends on current in individual power conductors. For practically all distribution circuits, the speed of operation depends on the ratio of the fault current to the expected full-load current; the higher this ratio, the faster the operation. Considering the requirements for co-ordinated protection, it will be clear that the objective is to have the fault current sufficiently high with respect to full-load current so that prompt de-energization is practicable, and to have the de-energization equipment such as to take advantage of this condition.

Length and Location of Exposure

The length and location of the exposure with reference to the substation has an important bearing on the performance of co-ordinated protection. The most critical point at which a fault may occur is that where the fault current is at a minimum compared to the rating of the fuse or setting of the circuit-breaker relay which protects that part of the power circuit, and consequently where the de-energization will be the slowest expected.

There will be other possible points of contact where the fault currents may be substantially higher, but in these cases the speed of operation of fuses or circuit breakers will be higher, and the total energy dissipated at the fault may be no greater than it would be with the lower fault current at the critical point.

Generally, detailed consideration of the factors affecting co-ordinated protection is given only in the case of new and unusual situations. Usually this only involves assuring that the power circuit will de-energize promptly with a fault at the most critical location. Where the telephone plant is well grounded, it is not necessary to take its impedance into account in these calculations. But once the facts have been developed by a particular power company and telephone company, agreement can be reached on an area basis, and further studies of specific cases falling within the area of agreement are not required.

Field Experience

A measure of the satisfactoriness of joint-use arrangements in any voltage range is the experience of the utilities from the standpoint of electrical accidents to personnel, and plant damage. An analysis of electrical accidents indicates that there is no significant difference between what has in the past been called normal joint use and so-called higher voltage joint use. It is interesting that most of the accidents occur in connection with construction activities and are of such nature that their occurrence cannot be related to voltage. The type of accidents which do occur unfortunately result too frequently from lack of attention to ordinary safety precautions and to failure to observe carefully thought out working practices.

Insofar as telephone plant damage is concerned, experience indicates that there is little relationship between the extent of the damage and the voltage on the joint-use lines. This is a result, to a large extent, from the more favorable de-energization characteristics of the higher voltage systems.

References

1. Power System Faults to Ground, Part I: Characteristics, C. L. Gilkeson, P. A. Jeanne, J. C. Davenport, Jr. *AIEE Transactions* (*Electrical Engineering*), vol. 56, April 1937, pp. 421–28.

2. Power System Faults to Ground, Part II: Fault Resistance, C. L. Gilkeson, P. A. Jeanne, E. F. Vaage. *Ibid.*, pp. 428–33.

3. Inductive Co-ordination of Common-Neutral Power-Distribution Systems and Telephone Circuits, J. O'R. Coleman, R. F. Davis. *Ibid.*, Jan. 1937, pp. 17–26, 189.

4. Protection Features for the Joint Use of Wood Poles, J. O'R. Coleman, A. H. Schirmer. *AIEE Transactions* (*Electrical Engineering*), vol. 57, March 1938, pp. 131–40.

5. Noise Co-ordination of Rural Power and Telephone Systems, H. W. Wahlquist, T. A. Taylor. *AIEE Transactions* (*Electrical Engineering*), Vol. 57, Nov. 1938, pp. 613–21.

6. Co-ordination of Power and Communication Circuits for Low-Frequency Induction, J. O'R. Coleman, H. M. Trueblood. *AIEE Transactions* (*Electrical Engineering*), vol. 59, July 1940, pp. 403–12.

7. Inductive Co-ordination of *REA* Distribution Systems and Telephone Systems, K. J. Plucknett, W. T. Smith, T. A. Taylor. *AIEE Transactions*, vol. 60, pp. 586–95.

8. Joint Pole Practices for Supply and Communication Circuits, Joint Committee on Plant Co-ordination. *Report No. M-12*, Edison Electric Institute, Bell Telephone System, New York, N. Y., 1945.

No Discussion

Type-ON Carrier Telephone

R. D. FRACASSI
NONMEMBER AIEE

H. KAHL
NONMEMBER AIEE

IN THE normal process of laying out toll circuits, it is found that many of the circuits will be made up of a combination of open wire and cable facilities. The new short-haul carrier systems will be found on many of the facilities, type-*O* on open wire[1] and type-*N* on cable.[2] Consequently, in built-up circuits, it will frequently be necessary to interconnect *O* and *N* carrier facilities at the open wire cable junction. The junctions of open wire and cable will occur most frequently at population centers which serve as terminals for the two facilities. However, there will be numerous cases where the junction will occur at the frontier of the cable expansion program, and may represent a complete change of facility from open wire to cable in the center of a route.

At a junction it is, of course, always possible to connect any two systems together on a channel basis by connecting terminals back-to-back at voice frequency. Thus, one 12-channel *N* system terminal can be used to extend the channels of three 4-channel *O* system terminals. This arrangement involves complete circuits in each of the terminals and is obviously an extravagant use of equipment. Some equipment saving can be made by the omission of the compandor and signaling circuits. These circuits are standard equipment with both type-*O* and type-*N* terminals, and when omitted they must be replaced by supplementary units which are called through-channel circuits and which are now in production. However, the interconnection of channels on any voice frequency basis is not attractive economically.

The greatest saving in equipment is effected when the transition from type-*O* facility to type-*N* facility is made at carrier frequencies. Such a transition is necessarily restricted in form and does not mean the conversion of the single sideband type-*O* signals to the double sideband type-*N* signals. The practical approach involves the conversion of the type-*O* signals, in groups of four channels, to frequency allocations capable of being transmitted over the *N* line.

It is feasible to transmit *O*-type signals over *N* repeatered cable pairs if essentially the same range of carrier frequencies is used. That is, if the *O* frequencies are translated into the *N* frequency range they can be transmitted over *N* facilities in place of the regular *N* transmission. This transmission of *O*-type signals over cable circuits equipped with *N* repeaters has been designated the *ON* system, that is *O* transmission over *N* facilities. To accomplish this transmission requires that the *O* transmission from the open wire line be transferred to the *N* carrier frequency range, and also that it must be co-ordinated in level and direction of transmission with the regular *N* systems which may also exist in the same cable. The transmission of type-*O* signals over a type-*N* line also forms the basis for a 20-channel all-cable system.

The *ON* system, like the *O* and *N* systems, is intended for short-haul use. A circuit in an *ON* system should provide satisfactory service for distances up to at least 200 miles, including the open wire portion of the circuit when type *O* is involved.

Two important considerations in the development of this equipment were:

1. To provide maximum flexibility in the translation of any *O* line frequency to any position in the *ON* band without the need of special equipment.
2. To make use of as much of the presently developed *O* and *N* equipment as possible, not only to keep the development of new equipment to a minimum, but also to make the *ON* system available at the earliest possible date.

Frequency Allocation

Fig. 1 shows the frequency allocation of the *ON* system. Also shown for comparison are the standard *N* line frequencies. It is seen that the low-frequency *N* group extends from 44 to 140 kc; in comparison, the *ON* low-groups range is from 40 to 136 kc and comprises the five standard *O* system line frequencies, namely the *OB* system low and high groups, the *OC* system low and high groups, and the *OD* system low group. The reason for using this frequency allocation is that these groupings correspond to the passbands of the type *O* components used in the type *ON* equipment. Although the *ON* low-groups range extends down to 40 kc, which is 4 kc below the standard low-group *N* range, those additional frequencies are satisfactorily transmitted by standard *N* repeaters.

Although the standard *O* equipment provides the allocation only in the low-group band of the type-*N* carrier line, this group by means of an *ON* repeater may be translated to a frequency range corresponding to the high *N* carrier line frequencies. Fig. 1 also shows the frequency allocations of the high *N* and the high *ON* frequency groupings. Since the *ON* frequency range fell 4 kc below the *N* in the low-group range, it follows that it will extend 4 kc above in the high-group, but these additional frequencies can also be transmitted by *N* repeaters.

It is observed that the *ON* groups, consisting of five standard type-*O* line groups, add up to a total of 20 channels. One particular difference exists between the grouping used for *ON* and the *O* line frequencies in that the individual 4-channel groups are arranged side by side for one direction of transmission in the *ON* arrangement, in contrast to the *O* line arrangement in which the 4-channel groups are oppositely directed. Furthermore in the 96-kc range of an *ON* group there are four 4-kc gaps which provide separation between adjacent *O* groups. This permits these groups to be separated by means of standard *O* filters.

Although the allocation of the *ON* frequencies in the *N* band does not make complete use of the frequency space available, the advantages may be summarized as follows:

1. The *ON* low-groups frequencies correspond to type-*O* line frequencies; thus standard *O* filters and frequency grouping can be used.
2. The separation of groups permits individual groups of four channels to be added as desired using standard *O* equipment.
3. Essentially standard *O* terminals can be used at the end of the cable portion of the circuit.
4. The flexibility of being able to translate any *O* line frequency group to any position in the *ON* group is attained.

System Description

The *ON* system is fundamentally an arrangement whereby five 4-channel

---

Paper 53-333, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Pacific General Meeting, Vancouver, B. C., Canada, September 1–4, 1953. Manuscript submitted May 26, 1953; made available for printing July 14, 1953.

R. D. FRACASSI is with the Bell Telephone Laboratories, Inc., Murray Hill, N. J., and H. KAHL is with the Bell Telephone Laboratories, Inc., New York, N. Y.

The authors wish to acknowledge the contributions made by other members of the technical staff of the Bell Telephone Laboratories in the development of the *ON* system; also the collaboration of members of the Department of Operation and Engineering of the American Telephone and Telegraph Company in outlining field requirements and the problems to be met in the technique of application.

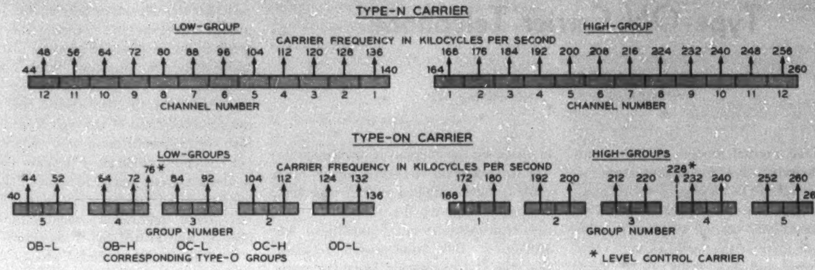

TYPE-N CARRIER
LOW-GROUP
HIGH-GROUP
CARRIER FREQUENCY IN KILOCYCLES PER SECOND
44 48 56 64 72 80 88 96 104 112 120 128 136 140
12 11 10 9 8 7 6 5 4 3 2 1
CHANNEL NUMBER
164 168 176 184 192 200 208 216 224 232 240 248 256 260
1 2 3 4 5 6 7 8 9 10 11 12
CHANNEL NUMBER
TYPE-ON CARRIER
LOW-GROUPS
HIGH-GROUPS
CARRIER FREQUENCY IN KILOCYCLES PER SECOND
40 44 52 64 72 76* 84 92 104 112 124 132 136
5 4 3 2 1
GROUP NUMBER
OB-L OB-H OC-L OC-H OD-L
CORRESPONDING TYPE-O GROUPS
168 172 180 192 200 212 220 228* 232 240 252 260 264
1 2 3 4 5
GROUP NUMBER
* LEVEL CONTROL CARRIER

Fig. 1. Type-ON carrier frequency allocation

type-*O* groups or systems are transmitted over a type-*N* line. Fig. 2 illustrates the general equipment layout for a typical 20-channel open-wire cable *ON* arrangement. The equipment involved at the junction of the open wire and cable includes essentially standard type-*O* units and a modified type-*N* repeater. The *ON* junction converts the five assigned type-*O* systems to the low-group frequency band of the type-*N* system. The *ON* repeater provides amplification, slope adjustment of carrier amplitudes, and frequency conversion to the type-*N* high-group band as required to co-ordinate with the standard type-*N* transmission. The *ON* terminal at the other end of the cable provides for the termination of 20 channels and again consists almost wholly of standard type-*O* units.

The *ON* system may also be used in all-cable layouts. The equipment requirements for such an arrangement providing for the transmission of 20 channels over a single *N* line are shown in Fig. 3.

Modulation Plan

The type-*ON* carrier modulation plan shown in Fig. 4 is designed to place five different 4-channel groups, each corresponding to a type-*O* group or system in a frequency band capable of being transmitted over a type-*N* line. The plan is based on the use of the basic 4-channel band of 180 to 196 kc provided by the type-*O* equipment. Each basic 4-channel group is converted to an allocated *ON* low-groups position by use of a group modulator and associated frequency, according to Fig. 4. The full *ON* complement of five groups covers the frequency band from 40 to 136 kc.

When the system layout requires that transmission to the first repeater section must be in the low-groups band in order to co-ordinate with type-*N* systems, the band of frequencies from 40 to 136 kc is applied to the line by the *ON* repeater without further modulation. When high-groups are required, the *ON* repeater applies a further step of modulation with a carrier frequency of 304 kc so that the *ON* low-groups are converted to *ON* high-groups. Modulation in the repeaters along the type-*N* line converts these bands of frequencies back and forth between low-groups and high-groups in a manner similar to the type-*N* plan.

The *ON* modulation plan may be considered a 2-step process when applied at the junction of open wire and cable. Four-channel groups are received on the open wire line side of the junction at type-*O* line frequencies and then each is modulated to the basic 4-channel band of 180 to 196 kc. The second step modulates each group to its *ON* low-groups allocation. Since the basic 4-channel carrier frequency band of 180 to 196 kc is common to each of the *OA*, *OB*, *OC* and *OD* systems, it follows that the five groups of *ON* may be composed of any combination of 4-channel type-*O* systems. This flexibility is an essential requirement for a practical system since it avoids undesirable restrictions on its field of application.

A simpler single-step modulation plan was considered for the translation from the type-*O* line frequencies to the *ON* frequencies but rejected because of the limitations involved. This plan made use of frequency frogging repeaters similar to those used in *OB*, *OC*, and *OD* carrier. In a type-*OB* repeater, for example, an input frequency band of 40 to 56 kc is converted to an output band of 60 to 76 kc. The same procedure could be used to translate a type-*O* 4-channel group of line frequencies to an allocation in the *ON* low-groups band by appropriate selection of modulating frequency and filters.

One of the factors which influenced the rejection of the single-step modulation plan was concerned with the large number of redesigned filter units that would be required. The standard type-*O* auxiliary filters used in repeaters contain a section with a low-group passband and another with a high-group passband. For use in a single-step modulation plan for *ON* there would be a need for a large number of redesigned auxiliary filters to provide for the numerous combinations of type-*O* line frequencies and *ON* low-groups allocations.

A second factor bearing on the choice of the 2-step modulation plan over the single-step arrangement was concerned

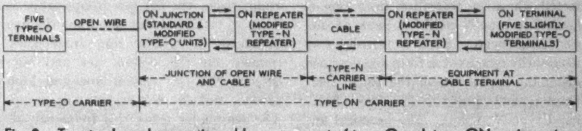

FIVE TYPE-O TERMINALS
OPEN WIRE
ON JUNCTION (STANDARD & MODIFIED TYPE-O UNITS)
ON REPEATER (MODIFIED TYPE-N REPEATER)
CABLE
ON REPEATER (MODIFIED TYPE-N REPEATER)
ON TERMINAL (FIVE SLIGHTLY MODIFIED TYPE-O TERMINALS)
JUNCTION OF OPEN WIRE AND CABLE
TYPE-N CARRIER LINE
EQUIPMENT AT CABLE TERMINAL
TYPE-O CARRIER
TYPE-ON CARRIER

Fig. 2. Twenty-channel open-wire cable arrangement of type-O and type-ON carrier system

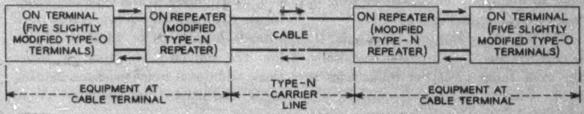

ON TERMINAL (FIVE SLIGHTLY MODIFIED TYPE-O TERMINALS)
ON REPEATER (MODIFIED TYPE-N REPEATER)
CABLE
ON REPEATER (MODIFIED TYPE-N REPEATER)
ON TERMINAL (FIVE SLIGHTLY MODIFIED TYPE-O TERMINALS)
EQUIPMENT AT CABLE TERMINAL
TYPE-N CARRIER LINE
EQUIPMENT AT CABLE TERMINAL

Fig. 3. Twenty-channel all-cable arrangement of type-ON carrier system

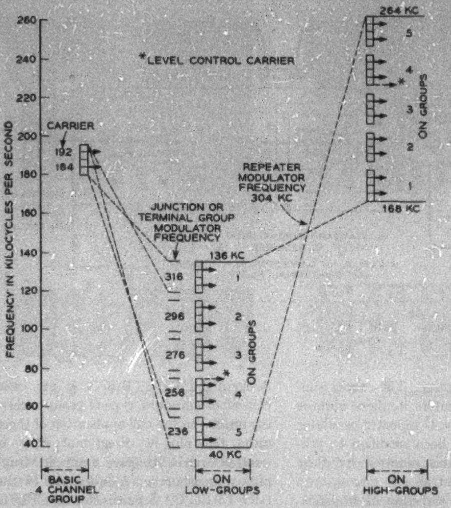


Fig. 4. Type-ON carrier modulation plan

with the provision for *OA* systems. The standard *OA* repeater is a nonfrequency-frogging circuit, and use of the single-step modulation plan would require the design of a new repeater circuit.

The *ON* modulation plan, as described in the preceding, applies to the transmitting direction and represents the modulation steps applied in going from the basic 4-channel band of the *ON* junction or terminal to the frequencies applied to the pair in the cable. In the receiving direction the process is reversed. Frequencies are received from the type-*N* line in the band opposite to that used on the transmitting side. When the frequencies arriving on the receiving cable pair are in the *ON* low-groups band the *ON* repeater supplies amplification without modulation. Selection and modulation in 4-channel groups is then carried out by the junction or terminal equipment in accordance with properly allocated filters and group oscillator frequencies.

When the received signals are in the *ON* high-groups band, the *ON* repeater supplies a step of modulation to convert these to the low-groups band for proper selection in 4-channel groups.

Junction

A block schematic of the junction equipment is shown in Fig. 5. The 20 channels to be transmitted over the *ON* system are shown to be composed of 16 channels from a complete type-*O* system and four channels from an *OB* system which is being transmitted over a pair of wires in the frequency spectrum just above a type-*C* system. A representative arrangement might equally well consist of 20 channels derived from five different *OB* systems.

The line filters involved in any arrangement of facilities are the same as

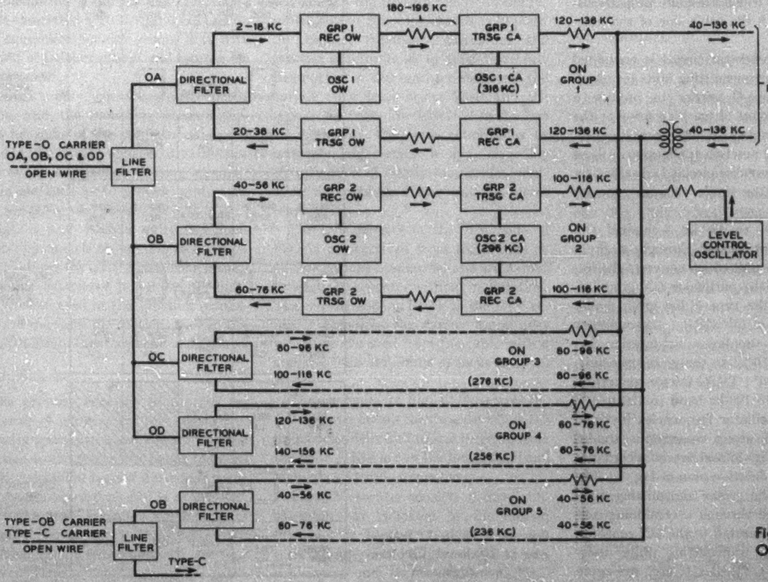

180–196 KC
2–18 KC
GRP 1 REC OW
GRP 1 TRSG CA
120–136 KC
40–136 KC
OA
DIRECTIONAL FILTER
OSC 1 OW
OSC 1 CA (316 KC)
ON GROUP 1
TO ON REPEATER
20–36 KC
GRP 1 TRSG OW
GRP 1 REC CA
120–136 KC
40–136 KC
TYPE–O CARRIER
OA, OB, OC & OD
OPEN WIRE
LINE FILTER
40–56 KC
GRP 2 REC OW
GRP 2 TRSG CA
100–116 KC
OB
DIRECTIONAL FILTER
OSC 2 OW
OSC 2 CA (296 KC)
ON GROUP 2
LEVEL CONTROL OSCILLATOR
60–76 KC
GRP 2 TRSG OW
GRP 2 REC CA
100–116 KC
80–96 KC
OC
DIRECTIONAL FILTER
ON GROUP 3
80–96 KC
100–116 KC
(276 KC)
80–96 KC
120–136 KC
OD
DIRECTIONAL FILTER
ON GROUP 4
60–76 KC
140–156 KC
(256 KC)
60–76 KC
40–56 KC
OB
DIRECTIONAL FILTER
ON GROUP 5
40–56 KC
TYPE–OB CARRIER
TYPE–C CARRIER
OPEN WIRE
LINE FILTER
60–76 KC
(236 KC)
40–56 KC
TYPE-C

Fig. 5. Type-ON carrier junction

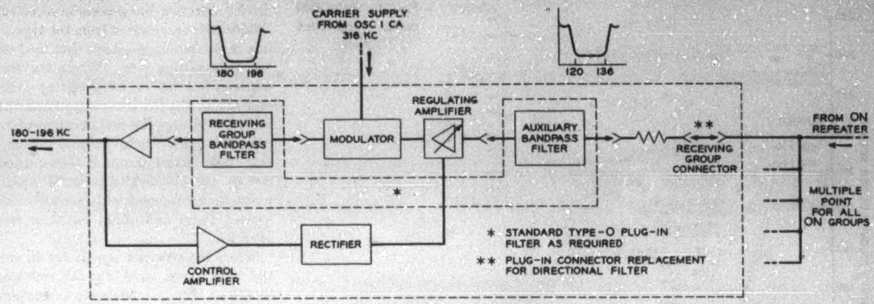

CARRIER SUPPLY FROM OSC 1 CA 316 KC
180 196
120 136
180-196 KC
RECEIVING GROUP BANDPASS FILTER
MODULATOR
REGULATING AMPLIFIER
AUXILIARY BANDPASS FILTER
**
RECEIVING GROUP CONNECTOR
FROM ON REPEATER
MULTIPLE POINT FOR ALL ON GROUPS
*
RECTIFIER
CONTROL AMPLIFIER
* STANDARD TYPE-O PLUG-IN FILTER AS REQUIRED
** PLUG-IN CONNECTOR REPLACEMENT FOR DIRECTIONAL FILTER

Fig. 6. Type-ON carrier. GRP 1 REC CA

those used in type-*O* and identical mounting arrangements are used.

The circuits used in the junction for the selection of each 4-channel type-*O* group and the transformation of each group from the line frequencies to the basic band of 180 to 196 kc are identical with those used in type-*O* terminals. These consist of group receiving, group transmitting and group oscillator circuits. In Fig. 5 they are shown designated GRP 1 REC OW, GRP 1 TRSG OW, OSC 1 OW, etc. The numeral designation serves to identify the *ON* group assignment and the OW suffix distinguishes these circuit units from others which are more properly associated with the cable side of the junction.

A noise generator circuit is contained in the group transmitting units for use in standard type-*O* carrier to provide a nominal amount of masking noise at the channel terminals. This feature is not required in the junction application of these units and the noise circuit is disabled by removal of the single vacuum tube involved.

Each 180- to 196-kc 4-channel *ON* group is next passed through a group transmitting unit which converts the frequencies to the particular *ON* group allocation in the type-*N* low-group band. In the case of *ON* GRP 1 the oscillator OSC 1 CA supplies a modulation frequency of 316 kc to the group transmitting unit GRP 1 TRSG CA for conversion of the 180- to 196-kc band to 120 to 136 kc. The oscillator frequencies involved in other *ON* group conversions are indicated in Fig. 5, and are in accordance with the modulation plan of Fig. 4. The outputs of the group transmitting units are multipled through a combining network and connected to the *ON* repeater.

The group transmitting units designated GRP 1 TRSG CA, etc., are again standard type-*O* units. The carrier supply for the modulators in these units is obtained from type-*O* repeater oscillator units which have been modified to provide the frequencies required by their application in type-*ON*.

In the opposite direction of transmission the signals received by the junction equipment from the *ON* repeater are also in the *ON* low-groups frequency band. Five group receiving circuits, one for each *ON* group, are used with their inputs arranged in multiple. These circuits are designated GRP 1 REC CA, etc., and consist of standard type-*O* group receiving units. In the *ON* application of these units the plug-in directional filter is omitted and in its place a simple strapping plug is used. A schematic of the GRP 1 REC CA circuit is shown in Fig. 6. The auxiliary filters used in these applications are standard type-*O* designs. Each *ON* group requires a different code to select properly the 4-channel group involved. In each *ON* group, carrier supply for the modulator in the GRP

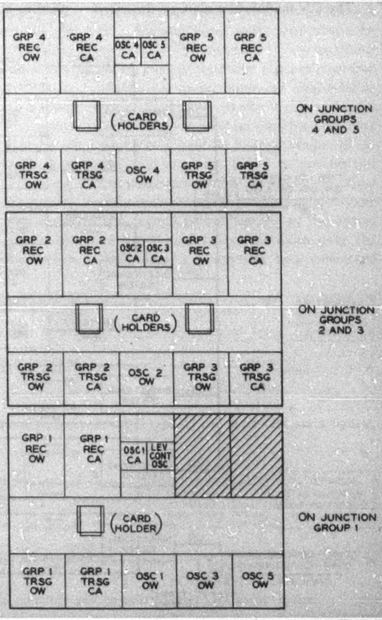

GRP 4 REC OW
GRP 4 REC CA
OSC 4 CA
OSC 5 CA
GRP 5 REC OW
GRP 5 REC CA
CARD HOLDERS
GRP 4 TRSG OW
GRP 4 TRSG CA
OSC 4 OW
GRP 5 TRSG OW
GRP 5 TRSG CA
ON JUNCTION GROUPS 4 AND 5
GRP 2 REC OW
GRP 2 REC CA
OSC 2 CA
OSC 3 CA
GRP 3 REC OW
GRP 3 REC CA
CARD HOLDERS
GRP 2 TRSG OW
GRP 2 TRSG CA
OSC 2 OW
GRP 3 TRSG OW
GRP 3 TRSG CA
ON JUNCTION GROUPS 2 AND 3
GRP 1 REC OW
GRP 1 REC CA
OSC 1 CA
LEV CONT OSC
CARD HOLDER
ON JUNCTION GROUP 1
GRP 1 TRSG OW
GRP 1 TRSG CA
OSC 1 OW
OSC 3 OW
OSC 5 OW

Fig. 7. Physical arrangement of 20-channel ON junction equipment

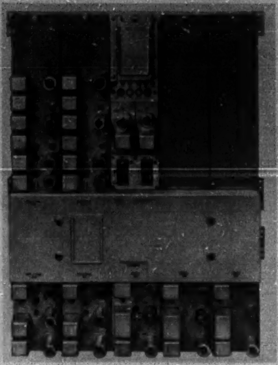


Fig. 8. ON junction group 1

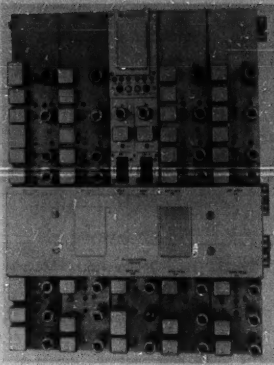


Fig. 9. ON junction groups 4 and 5

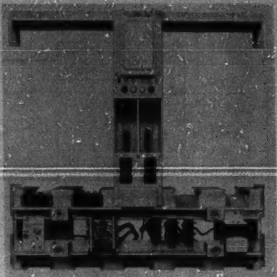


Fig. 10. Junction framework ON group 1

REC CA circuit is obtained from the same source that supplies the modulator in the GRP TRSG CA circuit.

The physical arrangement for a 20-channel *ON* junction is shown by the sketch in Fig. 7. The circuit units are mounted on three different junction mountings. The basic mounting is designated *ON* junction GRP 1, and it contains the combining networks and level control oscillator. It also provides overflow mounting space for the OW oscillators associated with *ON* groups 3 and 5. The equipment for *ON* groups 2 to 5 is divided in pairs of groups and located on two mountings.

Power supply circuits, fuses, alarm lamps, and the mounting accommodations for some of the oscillators are provided by the plug-in upper central units in each of the mountings. Fusing arrangements are such as to provide protection on a per-*ON*-group basis.

Figs. 8 and 9 show fully equipped laboratory models of junction mountings. Fig. 10 shows the GRP 1 mounting framework and associated power supply unit. The shelves used in the junction mountings are the same die castings used in the type-*O* terminal and repeater mountings.

ON REPEATER

The primary function of the *ON* repeater is to convert the low-groups frequencies transmitted and received by the junction to the frequency bands required by the associated type-*N* line. It is made up of a standard type-*N* LH or HL section and a new LL section which includes an equalizing network. These sections or subassemblies may be combined into four different arrangements to meet the needs of different applications. The optional arrangements are as shown in Table I.

The illustration in Fig. 11 shows a LH LL type-*ON* repeater schematic. The LH refers to the W-E side of the repeater which converts the low-groups frequencies at its input to high-groups at its output. The +7-decibel (db) slope network at the W-E input has a rising loss characteristic with frequency. It introduces a 7-db slope in the magnitude of the carriers received from the *ON* junction. At the repeater output the slope is such as to place the highest frequency carrier at the greatest amplitude as in type-*N* carrier. Regulation in the repeater stabilizes the output on a total power basis and the type-*ON* carriers are delivered to the line at essentially the same amplitudes as corresponding carriers in type-*N*. This co-ordination of carrier amplitudes minimizes the problems of crosstalk and maintenance.

The *ON* carriers in the E-W side of Fig. 11 are received from the *N* line at *ON* low-groups frequencies and are passed through the repeater without modulation. The −7-db slope network in the E-W repeater equalizes the carrier amplitudes which are received from the cable with a nominal 7-db amplitude slope as in type-*N*.

The schematic of Fig. 12 illustrates the *ON* repeater circuit for the case where the frequencies delivered to and received from the type-*N* line are in the low-groups and high-groups bands respectively.

The *ON* repeater consists of three subassemblies as indicated by the dashed lines of Figs. 11 and 12. The subassemblies are interconnected by plug and jack connectors and the assembled *ON* repeater is the same in appearance as the standard *N* unit. The repeater is mounted on repeater mounting brackets on shelves like those of type-*N*, and may be operated adjacent to type-*N* repeaters. The associated *ON* repeater mounting wiring on the side facing the cables is the same as for type-*N* repeaters. This provides for standard *N* arrangements of over-the-cable power supply circuits, span pads, and artificial lines as required. The repeater mounting wiring facing the *ON* junction includes the +7-db slope network.

The *ON* repeater is also used at the *ON* terminal to provide the same functions described for the junction.

ON TERMINAL

The equipment at an *ON* terminal includes an *ON* repeater, a level control oscillator, and combining network circuit and five slightly modified type-*O* terminals. Each 4-channel terminal terminates one *ON* group and the modified terminals are accordingly identified as *ON* Terminal GRP 1, *ON* Terminal GRP 2, etc. Fig. 13 is a block schematic of a terminal arrangement.

In type-*ON* carrier, as in type-*O*, the transmitted carrier is used at the receiving terminal for channel regulation and demodulation. Prior to its use for either of these functions, the received carrier is passed through a selective circuit to remove extraneous signals which would be a source of interference. The pick-off

Table I. Optional Arrangements for the Sections of an ON Repeater

Options	W-E	E-W
1	LH	LL
2	HL	LL
3	LL	LH
4	LL	HL

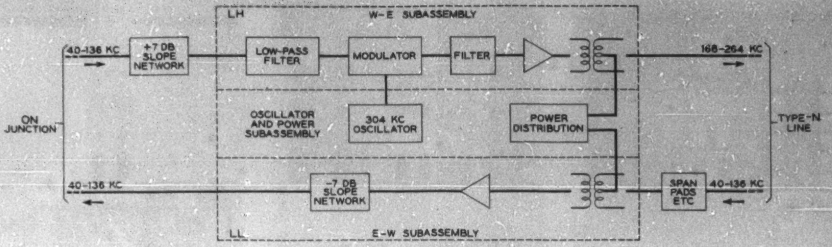

LH
W-E SUBASSEMBLY
40-136 KC
+7 DB SLOPE NETWORK
LOW-PASS FILTER
MODULATOR
FILTER
168-264 KC
ON JUNCTION
OSCILLATOR AND POWER SUBASSEMBLY
304 KC OSCILLATOR
POWER DISTRIBUTION
TYPE-N LINE
40-136 KC
-7 DB SLOPE NETWORK
SPAN PADS ETC
40-136 KC
LL
E-W SUBASSEMBLY

Fig. 11. Type-ON repeater LH LL unit

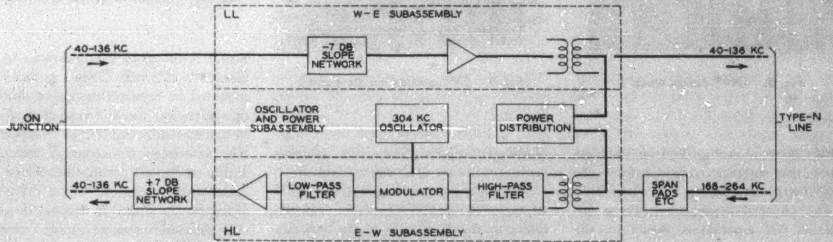

LL
W-E SUBASSEMBLY
40-136 KC
-7 DB SLOPE NETWORK
40-136 KC
ON JUNCTION
OSCILLATOR AND POWER SUBASSEMBLY
304 KC OSCILLATOR
POWER DISTRIBUTION
TYPE-N LINE
40-136 KC
+7 DB SLOPE NETWORK
LOW-PASS FILTER
MODULATOR
HIGH-PASS FILTER
SPAN PADS ETC
168-264 KC
HL
E-W SUBASSEMBLY

Fig. 12. Type-ON repeater LL LH unit

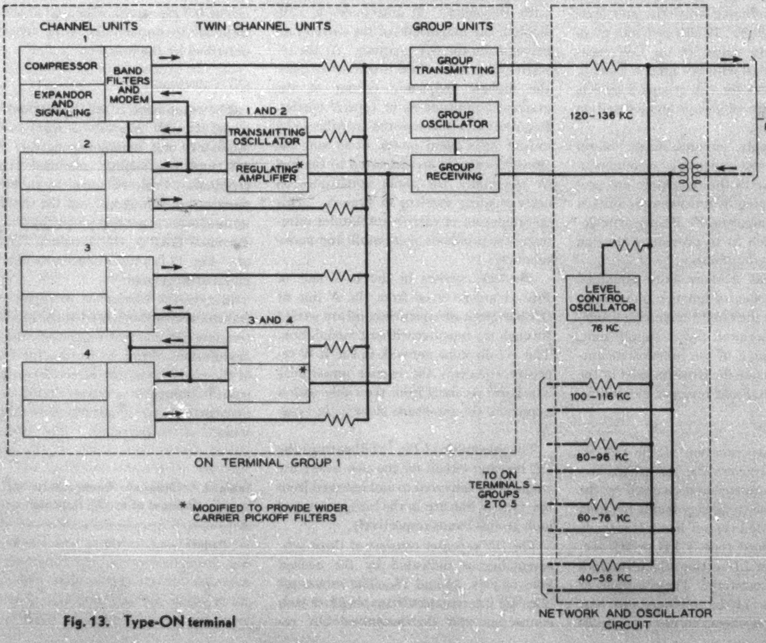
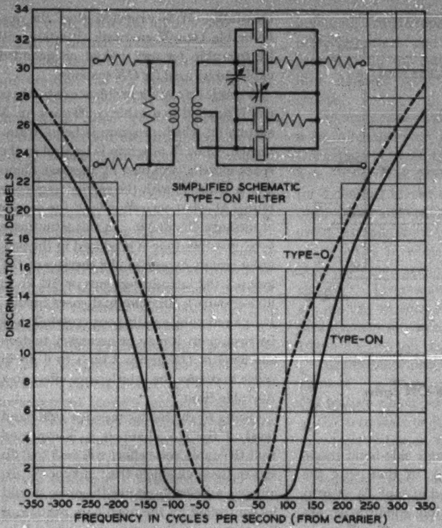

CHANNEL UNITS
TWIN CHANNEL UNITS
GROUP UNITS
1
COMPRESSOR
EXPANDOR AND SIGNALING
BAND FILTERS AND MODEM
GROUP TRANSMITTING
1 AND 2
TRANSMITTING OSCILLATOR
REGULATING AMPLIFIER*
GROUP OSCILLATOR
GROUP RECEIVING
120-136 KC
ON REPEATER
2
3
3 AND 4
*
4
LEVEL CONTROL OSCILLATOR
76 KC
100-116 KC
80-96 KC
60-76 KC
40-56 KC
TO ON TERMINALS GROUPS 2 TO 5
ON TERMINAL GROUP 1
* MODIFIED TO PROVIDE WIDER CARRIER PICKOFF FILTERS
NETWORK AND OSCILLATOR CIRCUIT

Fig. 13. Type-ON terminal



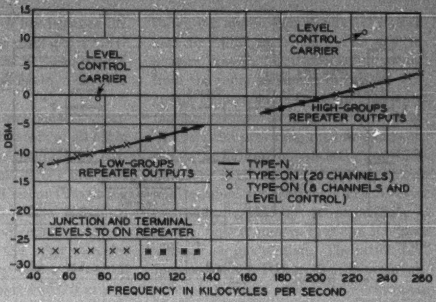


Fig. 14 (left). Type-ON carrier pick-off filter

Fig. 15 (above) Type-ON carrier levels

filter used for this purpose in type *O* has a passband of approximately ±50 cycles.

The required bandwidth of the pick-off filter is determined by the maximum error expected in the absolute frequency of the received carrier. Variations in frequency are contributed by the oscillators in the channel circuits, terminal group circuits, and frequency-frogging repeaters. In type-*O* the accumulated errors are not expected to exceed ±50 cycles for the longest systems.

In type-*ON* carrier the received carrier frequency error is expected to be greater than in type *O* because of the added variation introduced by the frequency-frogging repeaters of the type-*N* line. To care for this the pick-off filter has been redesigned to provide a passband of ±100 cycles. Twin channel units used in *ON* terminals or terminals of type-*O* systems involved in open wire-cable *ON* arrangements must be modified to provide the wider pick-off filter. The new filter is identical with the one it replaces from the standpoint of mounting arrangements and circuit connectons. Fig. 14 shows the loss characteristic of the new filter and its schematic circuit.

The directional filters have been omitted in the receiving group units and each 4-channel terminal is brought out on a 4-wire basis at carrier frequency as at the junction. The terminal group receiving circuits are similar to the one illustrated in Fig. 6 for the junction. In each *ON* group, the group oscillator frequency and the group receiving auxiliary filter are the same as those used at the junction. Masking noise is used at the terminals as in type *O*.

Partially Equipped *ON* Systems

When less than the full 20 channels are being transmitted over the *ON* system, the level control oscillators are used to provide an artificial load equivalent to the missing channel carriers. The purpose of the level control oscillators is to stabilize the amplitudes of the *ON* channel carriers in the *N* line at nominal values independent of the number of *ON* groups being transmitted. Without a level control carrier the effect of the *N* repeater regulator is to increase the magnitude of the remaining channel carriers as *ON* groups are dropped. The regulator circuit attempts to keep the total power at the repeater output constant.

The level control oscillator is used only when four or fewer *ON* groups are being transmitted over the *N* line. The recommended order of application of numbered *ON* groups is 1, 2, 3, 5, and 4. The level control oscillator output is reduced as system loading is increased, and it is removed altogether when group 4 is finally added.

Transmission Levels

The nominal amplitudes of the ten *ON* channel carriers in the type-*N* line are essentially the same as the corresponding type-*N* carriers. Fig. 15 shows the carrier amplitudes at low-groups and high-groups repeater outputs. The *ON* magnitudes fall on the sloped lines representing the positions of the type-*N* carriers. The chart also shows the carrier amplitudes delivered to and received from the *ON* repeater by the *ON* junction or terminal. The operating levels in the type-*O* units of the *ON* junction or terminal are the same as in standard type-*O* practice.

The carrier to side-band ratio of the *ON* channel circuits is unchanged from type-*O*. Both the message side-band level and the signaling tone amplitude are 6 db below the channel carrier.

Fig. 15 also illustrates the magnitude of level control carrier required to stabilize the channel carriers when only four carriers (eight channels) are being transmitted. As discussed in the preceding the level control carrier amplitude is lowered as channels are added, and removed completely when the full 10 *ON* carriers are applied.

Transmission Considerations

The over-all transmission performance of type-*ON* is expected to be comparable to type-*O*. The over-all frequency characteristic of an *ON* channel will be controlled largely by the type-*O* channel and group filters. When long type-*N* lines are involved, some degradation of the channel frequency characteristic will result from residual equalization effects in the line. These effects will be kept under control by the use of automatic deviation regulator circuits which have been developed recently for use on long type-*N* systems.

Phase distortion, an influence on the amplitude frequency characteristic of some channels in long type-*N* systems, is not a factor in type-*ON*. The single side-band *ON* type of transmission is not sub-

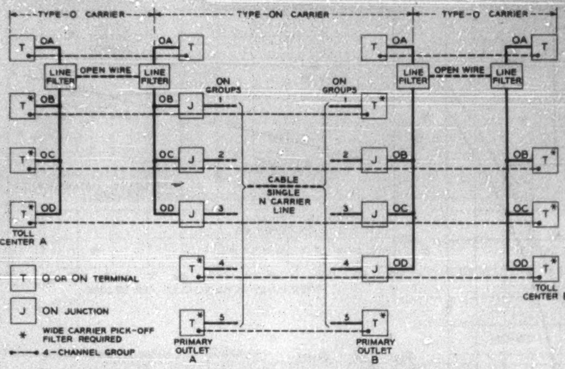

TYPE-O CARRIER
TYPE-ON CARRIER
TYPE-O CARRIER
T
OA
OB
OC
OD
LINE FILTER
OPEN WIRE
ON GROUPS
J
1
2
3
4
5
CABLE
SINGLE N CARRIER LINE
TOLL CENTER A
TOLL CENTER B
PRIMARY OUTLET A
PRIMARY OUTLET B
T O OR ON TERMINAL
J ON JUNCTION
* WIDE CARRIER PICK-OFF FILTER REQUIRED
4-CHANNEL GROUP

Fig. 16. Circuit layout illustrating flexibility provided by type-ON carrier

ject to amplitude effects such as those caused by the phase displacement of the two side bands used in type-*N*.

Channel net loss variations of *ON* channels are expected to be no greater than those in type-*O*. Two factors which are peculiar to *ON* operation affect transmission variations in opposite ways and should largely nullify each other. One factor, tending to improve channel net loss stability, involves the employment of extra regulating circuits in *ON* as compared to type *O*. For example, at the junction of open wire and cable, regulation is furnished by both the *ON* repeater and receiving group unit. Similarly, at the *ON* terminal the receiving circuit employes regulation in the *ON* repeater, group receiving circuit, and twin channel unit, as compared to only the group-receiving unit and twin channel unit in type-*O*.

The factor contributing towards a degradation of stability involves the individual channel regulators which are less effective with the single side-band transmission employed by *ON* than with the double side band used in type-*N*. Variations in the type-*N* line equalization are expected to produce noticeable effects in the channel net loss.

Owing to the difference in channel levels of type-*N* and type-*ON*, some inequality will exist in crosstalk between these systems. The double-side-band type-*N* system transmits each message side band at a level 15 db below the carrier. The single-side-band type-*ON* system operates with the side-band level 6 db below the carrier. With the carriers of type-*N* and type-*ON* being transmitted over the cable at like amplitudes this results in crosstalk from *ON* to *N* which is approximately 6 db greater than from another *N* system. Similarly the crosstalk from *N* to *ON* is approximately 9 db less than from another *ON* system. Only *ON* groups 2 and 4 have carriers which coincide with the type-*N* carriers and therefore offer the possibility of intelligible crosstalk. It is expected that the unfavorable type-*N* crosstalk situation will be tolerable for any paired or quadded cables to be used for *ON* systems.

A special crosstalk problem exists where *ON* is applied to a cable containing an *N* system on which program is being transmitted. The *N* program is located in the space usually occupied by message channels 5, 6 and 7, with the carrier frequency falling at what corresponds to the carrier of message channel 6. In this situation a 4-kc crosstalk tone is produced in the program circuit by one of the carriers of *ON* group 3 (92 kc in low-groups or 212 kc in high-groups.) In marginal cases involving paired cables with poor crosstalk loss between pairs it may be necessary to drop the level of *ON* group 3 by 3 to 6 db in order to reduce the interference effect to a tolerable limit.

Owing to the higher message side-band level of the *ON* channels it is expected that the cable noise effect will be 3 to 6 db less in these channels than in type-*N*. In an open wire-cable circuit the noise should be controlled by the open wire portion of the system, and performance should be comparable to type *O*.

Applications

The *ON* system, in its broader aspects, was originally planned as a part of the over-all type-*O* carrier development to provide flexibility and economy in short-haul circuit layouts involving different types of basic facilities. For example, the operation of *O* systems through long entrance or intermediate cables where the use of *N* instead of *O* repeaters would result in substantial economies; the operation of *O* carrier through cables carrying *N* systems with the ability to transfer to open wire facilities, on a carrier frequency basis, at intermediate points or cable terminals; as circuit growth and economic considerations permit, the use of *O* carrier on an open wire route to be developed as a

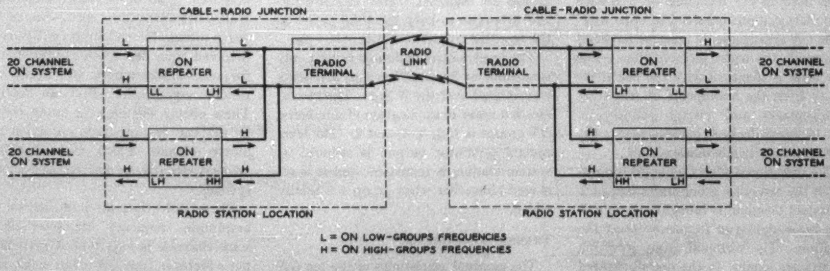

CABLE-RADIO JUNCTION
CABLE-RADIO JUNCTION
20 CHANNEL ON SYSTEM
20 CHANNEL ON SYSTEM
20 CHANNEL ON SYSTEM
20 CHANNEL ON SYSTEM
ON REPEATER
ON REPEATER
ON REPEATER
ON REPEATER
RADIO TERMINAL
RADIO LINK
RADIO TERMINAL
L
H
LL
LH
HH
RADIO STATION LOCATION
RADIO STATION LOCATION
L = ON LOW-GROUPS FREQUENCIES
H = ON HIGH-GROUPS FREQUENCIES

Fig. 17. Forty channels of ON carrier arranged for transmission over a radio link

carrier cable route, for all or a part of its length, by periodic cable extensions.

In addition to these applications, it now appears that the *ON* system will have a very considerable field of use as a 20-channel all-cable carrier system. A further important use anticipated for *ON* carrier is the future relief of *N* cables. This can be done by replacing *N* with *ON* terminals, thereby deriving eight additional channels per quad (or two pairs) without changes in the line facility and existing *N* repeaters. The *ON* system is also expected to find an important field of application in channelizing radio systems (up to 40 channels being obtainable) as outlined later in more detail.

A more detailed example of effective use of the *ON* equipment is illustrated in the simulated circuit layout on Fig. 16. The arrangement shows a section of cable between two primary outlets with open wire facilities extending from each of these outlets to toll centers. The facilities involved consist of a complete type-*O* system from the toll center *A* to the adjacent primary outlet, a type-*ON* system involving a single type-*N* line between the primary outlets, and a complete type-*O* system to the toll center *B*. The dashed lines in light print indicate 4-channel groups.

The arrangement shows one *ON* group used to furnish four channels of all-cable *ON* carrier between the two primary outlets. Two other groups of *ON* are used for 4-channel blocks of circuits involving open wire and cable. One 4-channel group extends from toll center *A* to primary outlet *B*, while the other provides channels from primary outlet *A* to toll center *B*. The remaining two *ON* groups are used in an open wire-cable–open wire arrangement providing eight channels between the two toll centers.

The flexibility provided by the arrangement of Fig. 16 is made possible by the *ON* feature which permits the application of *ON* terminals or junctions at the ends of a type-*N* line in any combination desired.

Type-*ON* also provides an effective method for channelizing circuits over a radio link. The arrangement in Fig. 17 shows the application of forty *ON* channels to radio terminals. The *ON* repeaters provide for transmission to and from the radio terminals in the *ON* low-groups frequency band of 40 to 136 kc, and the high-groups band of 168 to 264 kc. The repeater providing the low-groups input to the radio terminals is the same unit described in connection with the junction and terminal applications. The repeater providing high-groups frequencies is a similar type of unit in that one side does not employ modulation. The latter unit, incorporating a HH section, is not available at present. Slope networks associated with the repeaters provide amplitude correction so that all channels are applied to the radio terminals at equal levels.

References

1. Type-*O* Carrier Telephone, J. A. Coy, E. K. Van Tassel. *AIEE Transactions*, vol. 71, pt. I, Jan. 1953, pp. 428–37.

2. A New Telephone Carrier System for Medium-Haul Circuits, R. S. Caruthers, H. R. Huntley, W. E. Kahl, Ludwig Pedersen. *Electrical Engineering*, vol. 70, Aug. 1951, pp. 692–97.

No Discussion

The Influence of Magnetic Amplifier Circuitry Upon the Operating Hysteresis Loops

HAROLD W. LORD
FELLOW AIEE

THE scope of this paper will be limited to the steady-state characteristics of those high-performance magnetic amplifier circuits which are classed as single-phase self-saturated types. No attempt will be made to indicate when considerations presented herein are applicable to the 100-per-cent external feedback type. Only resistive loads will be considered.

While the several basic methods of analysis of self-saturated magnetic amplifiers available today are helpful in predicting an approximate control characteristic curve, they lack the precision required by many designers. Experimental data show that the same magnetic components will have somewhat different control characteristics when used in different circuits. This effect is present in circuits which are so similar that the presently available methods of analysis predict that their control characteristics would be identical.

Many of the methods of analysis include the concept that the self-saturated magnetic amplifier is a magnetomotive-force-controlled device. Other methods consider it as a voltage-controlled device during the period in which the flux level is being reset. Either or both of these concepts will be employed in this paper, depending upon the circuit under discussion.

In order to facilitate a mathematical analysis, the magnetization characteristics of the core usually are so simplified that the results of the analysis are meaningless when rectangular hysteresis loop materials are considered. No attempt will be made to provide quantitative information in this paper. The operation of the several circuits to be considered will be related qualitatively to the actual operating dynamic minor hysteresis loop. (The term "dynamic hysteresis loop" as used herein is defined as the relationship between magnetic induction **B** and apparent magnetizing force **H**. The apparent magnetizing force includes that required by the static hysteresis loop, taken over the same operating flux densities, plus all magnetizing forces required to overcome eddy current effects associated with the rate-of-change of flux required by the operating conditions.)

It is the author's hope that, by considering the relationships of operating core characteristics to circuit operating conditions, the reader will obtain:

1. A better understanding of the basic circuit operation.

Paper 53-311, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Pacific General Meeting, Vancouver, B. C., Canada, September 1–4, 1953. Manuscript submitted June 2, 1953; made available for printing July 22, 1953.

Harold W. Lord is with the General Electric Company, Schenectady, N. Y. The author wishes to acknowledge the help of Myron Egtvedt in obtaining the oscillograms and data used in the paper.

2. Some of the reasons for the effect of the circuit upon the control characteristic.

3. A better appreciation of the part played by the dynamic magnetization properties of the core in determining those characteristics.

Some Pertinent Magnetic Characteristics

The rapid expansion of interest in, and application of, magnetic amplifiers has been attributed in large part to the development of magnetic materials, such as Deltamax, having very rectangular d-c hysteresis loops. However, nearly all of the analytical methods published so far fail to relate the control characteristic to a truly rectangular loop. Those methods which do consider the hysteresis loop usually show that a self-saturated magnetic amplifier would have an infinite gain or trigger action between full output and minimum output if the hysteresis loop were assumed to have an extremely high, or infinite, maximum differential permeability. It is therefore customary to assume a nonrectangular parallelogram type of hysteresis loop in order to circumvent this difficulty. Deltamax not only has a very rectangular d-c hysteresis loop but exhibits rectangular dynamic hysteresis loops with extremely high maximum differential permeabilities (exceeding 10^6). Although rectangular loop materials, used in self-saturated magnetic amplifiers, do show some trigger action in the region of cutoff,[1] such trigger effects are not limited entirely to rectangular loop materials. Moreover, the slope of the major portion of the control characteristic does not correspond to that of the high differential permeability, being much less than that predicted by the analysis. It appears, therefore, that an important factor is being neglected in the analysis. I believe this neglected factor involves the proper consideration of the eddy-current effects which cause the operating dynamic hysteresis loops to be wider than the static loop.

In an earlier paper[2] it was shown that there was a reduction in the eddy-current effects at 60 cycles by a thickness reduction from 0.002 inch to 0.001 inch. This eddy-current reduction cannot be accounted for by skin effect. The metallurgists believe it is related to domain boundary movements, detailed considerations of which are beyond the scope of this paper. However, the fact that eddy-current effects are present in thin gauge material, at rates of change of flux ($d\mathbf{B}/dt$) which correspond to very low frequencies, can be demonstrated through the use of a hysteresigraph.[3]

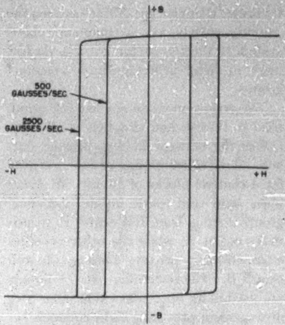


Fig. 1. Static loops of 65-Permalloy for two low rates of change of flux

The hysteresigraph automatically plots, on rectangular co-ordinates, the **B**-**H** curve of a test core. The curves of Fig. 1 were made on a machine which has sufficient long-time stability to allow several minutes for completing a curve and also has sufficient speed to permit one to be completed within a few seconds. The vertical portions of the two loops were made by setting the value of **H** to the indicated value and timing the change in induction **B** with a stopwatch. The test specimen was a spirally wound core of 0.005-inch 65-Permalloy. The narrowest loop was obtained at very low values of $d\mathbf{B}/dt$ and corresponds to the so-called d-c or static hysteresis loop of this core. A small increase of **H** in excess of that required by the static loop does not provide an instant or rapid change in flux. Instead, the $d\mathbf{B}/dt$ increases to such a value that the magnetomotive force in excess of that required by the static loop is balanced out by eddy currents associated with the domain boundary velocities required by the flux change. The loops show that for a greater excess of **H** the $d\mathbf{B}/dt$ is higher. It should be noted that for all of these loops the apparent maximum differential permeability ($d\mathbf{B}/d\mathbf{H}$) is infinite, since the flux is changing with a constant apparent **H**.

This dependence of $d\mathbf{B}/dt$ upon apparent **H** provides a means for explaining the observed performance of such rectangular hysteresis loop materials in those self-saturated magnetic amplifier circuits which have a high-impedance control circuit. (The term "high-impedance control circuit" as used herein will designate those control circuit conditions which do not allow the free flow of the even harmonic components of current in the control circuit. This is sometimes called "constrained magnetization."[4] The term "low-impedance control circuit" as used herein will designate those control circuit conditions which allow the free flow of the even harmonic components of current in the control circuit. This is sometimes called "natural magnetization."[4])

Ferrite magnetic materials have very high resistivities and therefore do not exhibit the previously described eddy-current effects at power and audio frequencies. If the metallurgists succeed in developing a ferrite having hysteresis loops whose maximum differential permeability is as high as that of Deltamax, then certain magnetic amplifiers employing such a material would be likely to have the triggering tendencies and performance characteristics as predicted by an analysis based upon its static loop.

Self-Saturating Single-Core Half-Wave Circuit

This circuit, shown in Fig. 2, is often analyzed as a basic element of the so-called doubler and full-wave circuits. The latter two circuits combine two similar cores which are assumed (sometimes erroneously) to operate into a common load resistance without affecting their individual control characteristics as obtained for the simple half-wave circuit. Since a single core provides no means for balancing out the fundamental components, the half-wave circuit must always have a relatively high impedance control circuit. This is, therefore, one circuit in which the current supplied by the control circuit is relatively constant at all instants throughout the cycle and so requires the eddy-current considerations in order to explain its operation when rectangular hysteresis loop core materials are employed.

Fig. 3(A) shows idealized static and dynamic major hysteresis loops and minor operating loops for a core material when used in this circuit; Figs. 3(B) and (C) show one cycle of operation of the circuit for minimum output and for firing angles of 120 degrees and 0 degrees (full output). The shaded areas emphasize the 120-degree firing angle conditions. The heavy curve in Fig. 3(B) indicates the induced voltage in the gate winding of the inductor, while the heavy curve in Fig. 3(C) indicates the current through the gate winding. In this case the rectifiers are assumed to be ideal in the inverse direction and the voltage drop across the load and rectifier due to the prefiring current is assumed to be almost negligible. Also, the applied voltage is made just sufficient to cause the peak-to-peak flux density swing at minimum output to be a fraction less

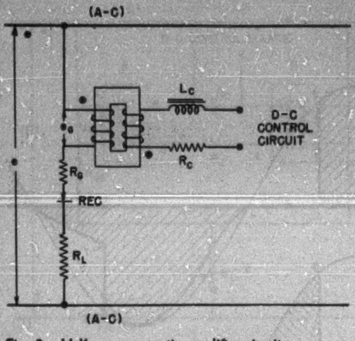


Fig. 2. Half-wave magnetic amplifier circuit

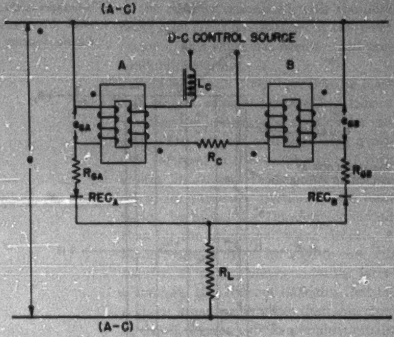


Fig. 4 (right). Doubler magnetic amplifier circuit with high-impedance control circuit

than two times the saturation density.

Consider first the condition of zero control current. After one cycle of operation following application of the alternating voltage, the core will remain in the region of $+B_s$ (swinging from $+B_r$ to 3) and the voltage across the coil will be substantially zero at all times, just as though it were a closed switch. The current through the gate winding will be half-cycles of sine waves as shown by the light solid curve of Fig. 3(C) and limited only by the total circuit and load resistance.

Now suppose the control current is adjusted to such a value as to provide a magnetomotive force corresponding to **H′** of Fig. 3(A). Then, during the negative half-cycle in which the rectifier is blocking, this value of **H′** will produce in the inductor core a negative volt-time area corresponding to the shaded portion in the negative half of Fig. 3(B). The flux density in the core will therefore be reset to **B′** during this period. When the next positive half-cycle occurs, the gate coil can absorb the positive volt-time area as shaded. This positive area is equal to the negative area provided during the preceding negative half-cycle, hence firing is delayed until 2′. The gate current during the prefiring period has a value corresponding to the total width of the minor hysteresis loop. This is true since the current in the gate winding not only must supply the current required by core excitation but, in addition, must balance out the magnetomotive force due to the control coil current. The control current continues to flow during the positive half-cycle and opposes the magnetomotive force required by the core during this period. After the core saturates, or fires, the current through the gate winding rises to 3′, being limited for the rest of the cycle only by the total circuit and load resistance.

If the control current is further increased to correspond to **H″**, the core will be completely reset to $-B_s$ during the negative half-cycle. Then, during the subsequent positive half-cycle, the core flux will just fall short of reaching saturation at $+B_s$ so the inductor will not fire. This is the condition of minimum output. For this condition the current through the gate coil during the positive half-cycles has a value corresponding to the total width of the dynamic hysteresis loop at

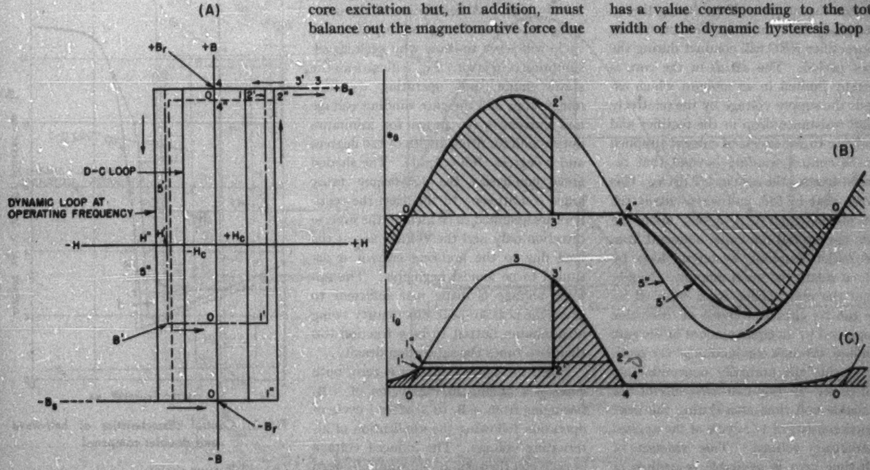


Fig. 3. Half-wave circuit operation

(A) Idealized hysteresis loops (B) Wave form of induced voltage in gate winding (C) Wave form of current in gate winding

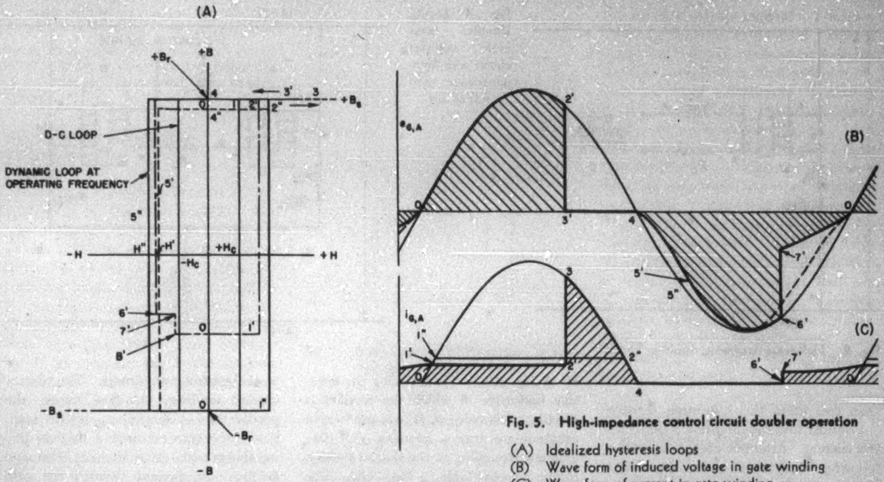


Fig. 5. High-impedance control circuit doubler operation

(A) Idealized hysteresis loops

(B) Wave form of induced voltage in gate winding

(C) Wave form of current in gate winding

the operating frequency and the control current required to establish this condition is determined by approximately one-half the width of this dynamic loop.

Consider now the effect of further increasing the control current so that it exceeds the value corresponding to **H"**. There will then be a tendency for the flux to reset to the $-\mathbf{B}_s$ value more rapidly than is required by the supply voltage. However, a greater $d\mathbf{B}/dt$ will tend to induce a voltage in the gate winding which exceeds the supply voltage during the reset half-cycle. Under such circumstances the rectifier *REC* will conduct during the reset period. The $d\mathbf{B}/dt$ in the core is thereby limited to an amount which exceeds the supply voltage by the relatively small resistance drop in the rectifier and load due to the excess of current supplied to the control winding beyond that required to reset the core in 1/2 cycle. This shows that in this half-wave circuit the net magnetizing force acting upon the core, can exceed the value obtained from the major dynamic hysteresis loop by only a relatively small amount. Nearly all of the magnetizing force supplied by the control circuit in excess of this value is opposed by an equal amount in the gate winding, through conduction of the rectifier during the normally nonconducting half-cycle. In this half-wave circuit the available volt-time area during the reset period consists of 1/2 cycle of the applied alternating voltage. This amount of volt-time area is available regardless of the firing angle during the forward or conducting half-cycle.

Doubler Circuit with High-Impedance Control Circuit

When two of the simple half-wave circuits are connected so as to supply alternating current to a common load resistor, as shown in Fig. 4, the resulting circuit is called the doubler circuit. This circuit is one which often has been erroneously assumed to have control characteristics similar to the simple half-wave circuit when the control circuit impedance is relatively high. A description of the operating conditions throughout a period of 1 cycle will serve to show why such an assumption is in error. Fig. 5 illustrates the static major and operating dynamic minor loops and the gate winding voltage and current wave forms for minimum output and for firing angles of 120 degrees and 0 degrees (full output). The shaded areas emphasize the 120-degree firing angle conditions. In this case the rectifiers are assumed to be ideal in the inverse direction only and the voltage across the load due to the prefiring current is assumed to be almost negligible. The applied voltage is made just sufficient to cause the peak-to-peak flux density swing at minimum output to be a fraction less than two times the saturation density.

When the control current is zero, both cores will remain in the region of $+\mathbf{B}_s$ (swinging from $+\mathbf{B}_r$ to 3) after 1 cycle of operation following the application of alternating voltage. The induced voltage in each coil then will be substantially zero at all times. The current through each gate winding will be half-cycles of sine waves as shown by the light solid curve of Fig. 5(C), one carrying current for half-cycles of one polarity and the other for half-cycles of the opposite polarity. The load current therefore will be full cycles of alternating current, the combined circuit in series with the load resistance acting as though it were a closed switch. The current will be limited only by the sum of the load resistance and the total

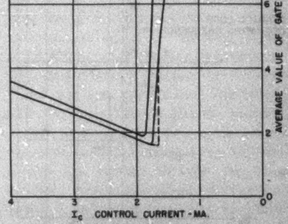


Fig. 6. Control characteristics of half-wave and doubler compared

(A) Half-wave circuit

(B) Doubler circuit with high-impedance control circuit

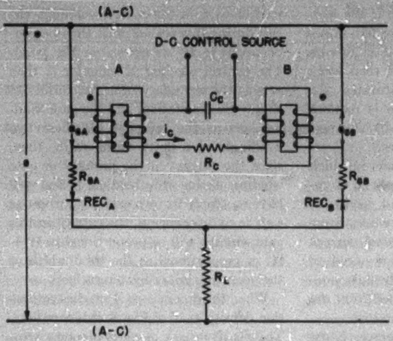


Fig. 7. Doubler magnetic amplifier circuit with low-impedance control circuit

resistance of one gate winding circuit. This latter resistance includes the forward resistance of one rectifier and the resistance of one gate winding.

Now suppose the control current is adjusted to such a value as to provide a magnetomotive force corresponding to **H**′ of Fig. 5. If we consider only the final steady-state conditions for the time being, then, during the half-cycle in which a rectifier is blocking, this value of **H**′ will produce in the inductor core associated with this rectifier a negative volt-time area corresponding to the shaded portion in the negative half of Fig. 5(B). The flux density in the core will therefore be reset to **B**′ during this period. This negative volt-time area curve has a large nick taken out of it, starting at a point in this half-cycle at which the other core fires, and applies a large part of the line voltage across the common load resistance. For the remainder of this negative half-cycle the negative volt-time area available for reset is no longer the full supply voltage but is the relatively small amount determined by the voltage drop across R_a of the branch circuit delivering current to the load. The net magnetomotive force required by the lower $d\mathbf{B}/dt$ for the remainder of the reset period is shown as 7′ on Fig. 5(A). The excess of control current, above this value, forces conduction of the rectifier associated with the core being reset for the remainder of this reset period. This effect has been termed preconduction current.

Suppose that a given core is to be reset to a certain value of average flux density in either the half-wave or the doubler circuit. The reduced volt-time area available for reset in the doubler circuit then makes it necessary that the average $d\mathbf{B}/dt$ for the early portion of the reset period be higher for the doubler circuit than for the simple half-wave circuit. A higher value of $d\mathbf{B}/dt$ for a given core will require a higher magnetomotive force. Therefore, this concept indicates that for a given firing angle, other than one close to cutoff, a given core in a doubler circuit with high control circuit impedance should require higher values of magnetomotive force than it does in a half-wave circuit. This is in agreement with the observed results obtained when a core is first tested in a half-wave circuit and then combined with a similar core in a doubler circuit having a high-impedance control circuit. Fig. 6 shows the results of such a test.

Doubler Circuit with Low-Impedance Control Circuit

Another author[5] has recently presented an excellent and extensive mathematical analysis, of which those portions describing the steady-state conditions are applicable to this circuit. The analysis there

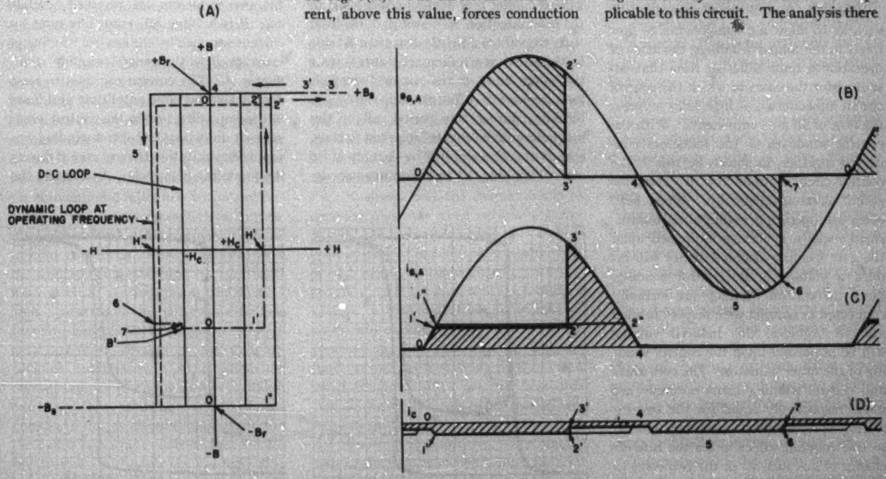


Fig. 8. Low-impedance control circuit doubler operation

(A) Idealized hysteresis loops
(B) Wave forms of induced voltage in gate winding
(C) Wave form of current in gate winding
(D) Wave form of current in control winding

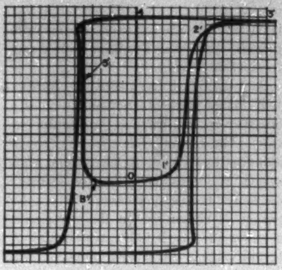


Fig. 9. Major and operating minor hysteresis loops for the half-wave circuit

presented is based upon a rectangular dynamic hysteresis loop but assumes that the control current is zero during the firing period. The description of the circuit operation which follows will include the effect of the static loop upon the control characteristic curve and the wave form of the current flowing in the control circuit. All postsaturation inductive effects will be neglected.

Fig. 7 shows schematically a doubler circuit with a control circuit made to have a low impedance to the flow of alternating currents by means of a capacitor C_c across the control source. If the control source has an inherently low impedance relative to the control winding resistance R_c, then capacitor C_c is not required.

When the control circuit impedance is such as to allow a-c components to flow freely in the control windings, the mode of operation is quite different from that described for the cases in which the control circuit impedance is so high as to suppress the flow of all a-c components. With the control windings of the inductors connected together, as shown by Fig. 7, the control circuit couples the two cores together electromagnetically. When both cores are operating within their unsaturated regions, a voltage impressed upon the gate winding of one inductor will induce a voltage in its control winding. Since the control windings are virtually connected in parallel with respect to such induced voltages, this induced voltage will be impressed upon the control winding of the other inductor. The core areas and control winding turns normally are made respectively equal for the two inductors of a particular magnetic amplifier, so this coupling effect forces the rate-of-change of flux ($d\mathbf{B}/dt$) in the two cores to be substantially the same at any given instant. The current which flows in the control circuit at any instant must therefore be such as to maintain this equality of $d\mathbf{B}/dt$ for the two cores. If the hysteresis loop is assumed to have zero slopes in the saturated regions and if all air-core inductive effects are neglected, then, when either one of the cores saturates, the $d\mathbf{B}/dt$ in the saturated core is reduced substantially to zero and no $d\mathbf{B}/dt$ is permitted in the unsaturated core.

An explanation of the manner in which this circuit operates is simplified if resistance effects are assumed negligible and if a particular steady-state firing angle be assumed. The control current amplitude and wave shape required to maintain the assumed steady-state condition then may be determined from the magnetic characteristics of the core.

If a firing angle of 120 degrees is assumed, then at the start of a half-cycle, when core A is to hold off by absorbing the applied voltage, the flux in this core will be at $\mathbf{B}'$ and point 0 on the operating minor hysteresis loop shown by Fig. 8(A). (Note that the wider loop with solid lines is the major dynamic hysteresis loop for the core at the operating frequency. The d-c hysteresis loop is the narrower loop shown with solid lines.) During the portion of the cycle that core A is holding off, the flux density in this core changes from $\mathbf{B}'$ to $+\mathbf{B}_s$ at point 2′ along the dot-dashed line. Figs. 8(B), (C), and (D) show the corresponding winding voltages and currents for inductor A. However, during this same period, core B, which has just completed its firing period, must have a similar flux density change but in the reset direction, so it swings down from $+\mathbf{B}_r$ to point 6 (slightly less than $\mathbf{B}'$ due to the effect of winding resistances) over the outer knee of the major hysteresis loop for core B. Since rectifier REC_B is blocking during this period, all of the magnetomotive force $\mathbf{H}''$ required to reset core B must be supplied by current in its control winding. The instantaneous current flowing in the control windings circuit during this period must thus be proportional to $\mathbf{H}''$ as shown in Fig. 8(A). The control winding of inductor A then acts as the secondary of a transformer to supply this excitation, and the gate winding acts as the primary. If these two windings have unity turns ratio, then, since the current in the primary or gate winding during this hold-off period will have to supply its own exciting current as well as that for core B, the current in this gate winding will be proportional to $\mathbf{H}'+\mathbf{H}''$ or proportional to the total width of the operating minor hysteresis loop.

When the flux in core A reaches saturation density at 2′ of Fig. 8, this core fires. The $d\mathbf{B}/dt$ in core A then becomes zero, so no voltage will be induced in its control winding. Since the average voltage across capacitor C is only that required to overcome the relatively low resistance of the control circuit for the average value of control circuit current, the voltage now impressed upon the control winding of inductor B is relatively small, hence the flux in core B, which has been reset to a value corresponding to point 6, continues to reset at a very low rate to $\mathbf{B}'$ during the conduction period of core A. Since the $d\mathbf{B}/dt$ in core B is very low, the current in the control circuit drops to substantially that corresponding to the static loop during this period as shown at 3′ in Fig. 8(D).

During the next half-cycle the roles of the two inductors are reversed. While core B is holding off, core A is reset by current supplied by the control winding of inductor B to the control winding of inductor A. The current required to reset core A has the same amplitude and flows in the same direction in the control winding as it did when the control winding was supplying excitation to reset core B during the preceding half-cycle. As a result, the

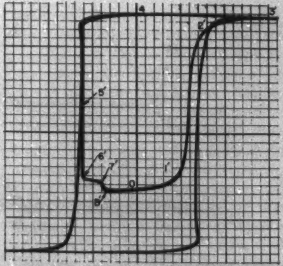


Fig. 10. Major and operating minor hysteresis loops for high-impedance control circuit doubler

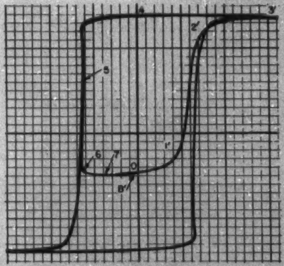


Fig. 11. Major and operating minor hysteresis loops for low-impedance control circuit doubler

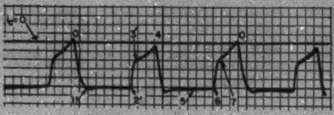


Fig. 12. Wave form of current in low-impedance control circuit of doubler

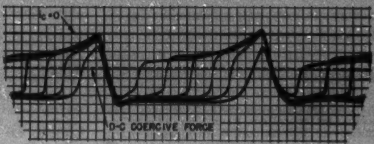


Fig. 13 (right). Composite oscillogram of current in low-impedance control circuit of doubler

current which flows in the control circuit during this half-cycle is as shown by points 4 through 7 to 0 in Fig. 8(D).

The effect of including a static hysteresis loop of finite width in the analysis of reference 5 is to shear the top of the control characteristic curve to the left by an amount which is proportional to the coercive force of the static loop. The values of load current and control current at cut-off would be unaffected. The gain would be a function of the difference between the dynamic coercive force and static coercive force.

Validity Checks

The operating minor hysteresis loops and current and voltage waveforms of experimental models of each of the magnetic amplifier circuits described in the foregoing were viewed on an oscilloscope screen. Vacuum-tube rectifiers were used for the experimental checks. The same specimen core of 0.002-inch Deltamax was used for all of these tests. Good correspondence was had between the idealized loops and experimental loops for cases employing a high-impedance control circuit.

Fig. 9 shows the major dynamic hysteresis loop and the operating minor hysteresis loop obtained for the simple half-wave circuit of Fig. 2 when firing at approximately 120 degrees. Similar shaped loops were obtained when this core was combined with a similar core to form the conventional full-wave d-c output circuits. The control characteristic curves, when normalized to an equivalent single-core basis, were practically identical with the control characteristic for the half-wave circuit. An analysis of the steady-state operation of the half-wave circuit is therefore applicable to single-phase magnetic amplifier circuits of the high-impedance control-circuit type when the normal output of the amplifier is direct current. The normal output of the doubler circuit is alternating current and the half-wave circuit analysis does not apply to it even though the a-c output is subsequently rectified.

Fig. 10 shows the major and operating minor hysteresis loops obtained for the doubler circuit of Fig. 4 when firing occurs at approximately 120 degrees. These should be compared with Fig. 5. The static coercive force of this specimen core was approximately 0.09 oersted and the dynamic coercive force at 60 cycles was approximately 0.17 oersted, a ratio of 1/1.9. The ratio of magnetomotive forces indicated between points 7′ and **B**′ to that at 5′ as shown by the operating minor loop of Fig. 10 is 1/1.5. The lower ratio of the observed value is probably due to the slight widening, over the static loop, required in order to provide the small $d\mathbf{B}/dt$ called for during the period from 7′ to **B**′.

When the doubler circuit of Fig. 7 was tested, the operating hysteresis loops were as shown by Fig. 11 for a 120-degree firing angle. Note that the bottom of the operating minor loops do not have the nick as was predicted by the assumed idealized conditions. Also, Fig. 12 shows that the control current during the conduction period of the gate winding does not remain constant as was predicted by the assumed idealized conditions. Both of these instances of poor correspondence between the predicted and actual results are due to the fact that postsaturation inductive effects were neglected. By neglecting this inductance, the voltage induced in the control coil of the saturated core was assumed to be zero during the saturation period. The control voltage was then assumed to be the only voltage acting upon the control circuit during the saturation period of each half-cycle. Fig. 12 shows that the current tends to decrease and may even reverse polarity during the saturation period. This is good evidence that during the latter half of each half-cycle the decreasing load current flowing in the gate winding induces a voltage in the control circuit which opposes the signal voltage and may be greater than the signal voltage. This actually causes the core being reset to lose a small amount of its maximum reset flux change during the saturation period, instead of gaining some as was predicted by the idealized assumptions. The amplitude of the instantaneous voltage induced in the control coil on the saturated core is a function of the slope of the **B-H** curve above the knee of the curve and of the instantaneous rate-of-change of load current. The amount of current which flows in the control circuit during this period is therefore dependent not only upon the control voltage and width of the static hysteresis loop but also upon the saturation-region inductance, load current, and inductance of the unsaturated control coil during this period. This latter inductance may be relatively low. The differential permeability of the core across the bottom of the minor hysteresis loop of rectangular loop materials is shown by Fig. 11 to be rather low. The unsaturated core is operating in this region when the saturated core is operating above the knee of its hysteresis loop.

It appears, therefore, that a rigorous analysis of self-saturated types of circuits having low control circuit impedances should include the effects of saturation-region inductance. This would include both the equivalent air-core inductance and the slope of the **B-H** curve above the knee of the curve. The differential permeability across the bottom of the minor loop also becomes an important factor to include in such an analysis. However, in spite of these limitations, the analysis of reference[5] has been shown to yield reasonably accurate results for materials such as grain-oriented silicon and overannealed Deltamax.

When the static hysteresis loop width of a very rectangular loop core material exceeds approximately 10 per cent of the dynamic loop width, a more rigorous analysis is required in order to predict with reasonable accuracy the control characteristic curve of magnetic amplifiers using such materials. Fig. 13 is a composite oscillogram of the control circuit current wave form, for representative Deltamax cores in the circuit of Fig. 7, when firing occurs at several different angles of delay. This indicates that an analysis which will predict the average value of current for each firing angle probably will be very difficult to achieve. Fig. 13 also demonstrates that both the dynamic and static hysteresis loops will be important factors in such an analysis.

Observations of the steady-state operating minor loops for other single-phase circuits of the low-impedance control-circuit type, such as the full-wave d-c output circuits, show only very minor deviations from those indicated by Fig. 11 for the doubler circuit.

References

1. An Instability of Self-Saturating Magnetic Amplifiers Using Rectangular Loop Core Materials, **S. B. Batdorf, W. N. Johnson.** *AIEE Transactions*, vol. 72, pt. I, July 1953, pp. 223–28.

2. Dynamic Hysteresis Loops of Several Core Materials Employed in Magnetic Amplifiers, **Harold W. Lord.** *AIEE Transactions*, vol. 72, pt. I, March 1953, pp. 85–88.

3. A Recording Fluxmeter of High Accuracy and Sensitivity, **P. P. Cioffi.** *Review of Scientific Instruments*, New York, N. Y., vol. 21, no. 7, 1950, pp. 624–28.

4. The Transductor, **Harold B. Rex.** *Instruments*, Pittsburgh, Pa., vol. 20, 1947, pp. 1102–09.

5. Theory of Magnetic Amplifiers with Square-Loop Core Materials, **H. F. Storm.** *AIEE Transactions*, vol. 72, pt. I, Nov. 1953, pp. 629–40.

◆

No Discussion

Flux Preset High-Speed Magnetic Amplifiers

C. B. HOUSE
ASSOCIATE MEMBER AIEE

EARLY in 1951, R. A. Ramey prepared a paper[1] on magnetic amplifiers. In this and succeeding reports[1,2] on the same subject the theory and application of a new type of half-cycle response magnetic amplifier were developed. This amplifier depended upon square hysteresis loop magnetic materials for its proper operation, and the analysis in turn postulated square-loop materials.

Further development in the Naval Research Laboratory has resulted in the circuit being presented here. This circuit extends the benefits of the half-cycle response to circuits utilizing magnetic materials not possessing the square-loop characteristics, i.e., the ratio of remanent flux to saturation flux may possess any value between 1.0 and 0.0. The circuit therefore, removes the restriction upon magnetic materials heretofore existing, and broadens the range of application of the half-cycle response circuits to include such materials as grain-oriented silicon steels and molybdenum-nickel-iron alloys, which have previously been deemed unsuitable. As an important corollary benefit, the circuit will overcome the deleterious effects suffered in the half-cycle response circuits with square-loop materials when using rectifiers which do not possess perfect rectifying characteristics.

Background

In reference 1, Ramey discussed in detail the theory of operation and circuit details of an amplifier which inherently possesses high response speeds (essentially one half-cycle of the power supply voltage) and still permits large power gains. The complete development of that theory is not within the province of this paper but a brief resumé will be given here to provide a basis from which the reader may follow the further developments.

The basic half-wave form of the Ramey amplifier is shown in Fig. 1. The conventions used in this diagram are as given in the following:

The polarities shown for the voltage sources e_{ac} and e_{ac}' are instantaneous values.

The polarities of the windings are as indicated.

The rectangular symbol between the windings indicates that the core is composed of a magnetic material with a rectangular hysteresis loop, as shown in idealized form in Fig. 2.

The turns ratio may be any number desired, as long as the voltages e_{ac} and e_{ac}' possess the same ratio in the same order.

The voltage e_c is an unidirectional signal voltage.

In the first section of the analysis it will be assumed that e_c is zero and that the control terminals are short-circuited together. The action of the circuit will then be followed until a steady-state condition results.

In beginning a step-by-step analysis, first it may be assumed that at a time when the applied voltage wave e_{ac}, e_{ac}' is going through zero (Fig. 3, point 1) the core will possess a remanent flux of a polarity and value determined by the previous history of applied voltages. To simplify discussion it is further assumed that the core had been previously saturated in the plus direction and at time of zero applied voltage the flux level is resetting at point 1 in Fig. 2. (Further justification of this assumption is contained in reference 1.)

As e_{ac} and e_{ac}' increase with the polarities shown in Fig. 1, the current in the left-hand or control circuit will produce a magnetomotive force equal to one-half the hysteresis loop width. The applied voltage will then react upon the core and change its state (or demagnetize it). The amount of change will be a function of the applied voltage and in usual design, 1/2 cycle of applied e_{ac}' will carry the flux level to just below saturation in the negative direction (Fig. 2, point 2). During the period just discussed, the applied voltage e_{ac} on the right-hand, or load side, will be directed against the rectifier, and no current will flow on that side. This half-cycle is commonly termed the reset cycle.

The next half-cycle (which will be called the gating half-cycle) starts at point 2 of Fig. 3. At this time, the voltage polarities, shown in Fig. 1, will reverse, so that the voltage applied to the control side will be blocked by the rectifier. However, on the load side, magnetizing current will start flowing and the applied voltage will now start magnetizing the core in the plus direction. With the assumed values of voltage, the flux will be carried just to the knee in the positive direction by the end of this half-cycle of applied voltage and thus return to point 1 of Fig. 2. A steady-state condition has now been reached and this sequence of operation will continue as long as voltage is applied. Only magnetizing current will flow through the load, and the amplifier will be in the state of minimum output with the flux swinging from knee to knee

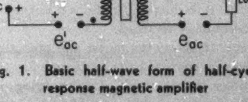


Fig. 1. Basic half-wave form of half-cycle response magnetic amplifier

---

Paper 53-337, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Pacific General Meeting, Vancouver, B. C., Canada, September 1–4, 1953. Manuscript submitted June 1, 1953; made available for printing July 9, 1953.

C. B. House is with the Naval Research Laboratory, Washington, D. C.

The help and guidance of D. G. Scorgie in the development of this material is gratefully acknowledged.

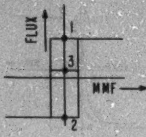


Fig. 2. Idealized hysteresis curve of a square-loop magnetic material

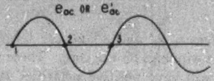


Fig. 3. Voltage wave of supply to magnetic amplifier

with no current greater than magnetizing current flowing.

Referring again to Fig. 1, it may be deduced that, during the half-cycle when e_{ac}' is applied through the control side, the total voltage applied to the core is the difference between e_{ac}' and e_c. If e_c is equal to zero, as has been previously assumed, the operation will be as explained. When e_c equals some finite value less than e_{ac}' the total voltage applied to the core on reset will be less than the voltage applied to the core in the previous half-cycle, and a new set of conditions will arise.

Under the new conditions the starting point for the analysis may remain the same; i.e., the applied voltage wave has gone through zero and the core flux is at position 1, on the hysteresis curve, Fig. 2, and the polarity of the applied voltage will become as shown in Fig. 1 in the next instant. The reset cycle is just beginning.

Since the time integral of the voltage applied to the core during the reset half-cycle will be less than the time integral of e_{ac}' the flux level will not be carried to point 2 on the hysteresis curve but will be changed only to some point part way down the loop (point 3) as determined by the net voltage time integral.

On the gating half-cycle the core will be able to absorb a voltage integral equal only to the net voltage integral applied on the reset cycle before it saturates and its impedance drops to zero. The load will then absorb the remaining voltage in the half-cycle, with current limited only by its impedance.

At point 3, Fig. 3, when the reset cycle starts, the core will again be at point 1, Fig. 2, and the cycle will repeat as before. The wave shape of the current through the load will be shown in Fig. 4.

Thus a magnetic amplifier has been described whose output power can respond to input signals within a half-cycle of the supply voltage. The time in the gating cycle at which the core saturates (or the firing angle) will be determined only by the net voltage time integral applied to the core on the reset half-cycle. If $\int_0^T (e_{ac}' - e_c)dt = 0$, or a negative number, the core will not reset at all, and full output should be realized. It readily follows that the control element need not be a voltage, but that any device which controls the net voltage applied to the core on the reset half-cycle may be used.[2]

The foregoing discussion has assumed perfect square-loop core materials and perfect rectifiers. However, complications arise when magnetic materials which do not have a perfectly square hysteresis loop, and rectifiers which have back leakage, are used. L. J. Johnson has discussed these effects.[3] In his attack upon the problem he established a theoretically ideal transfer characteristic curve and compared this ideal curve with the actual curves obtained when using different core materials in otherwise similar magnetic amplifier circuits. In his discussion of the suitability of the different magnetic materials, he made the comparisons of curve shape and cited possible causes of the deviations from the ideal curve. Among the causes were the effects of low remanence and rectifier back leakage.

Low remanence permits the core to reset partially without regard to the control voltage. This uncontrolled reset will result in the core-absorbing part of the voltage integral during the gating half-cycle, and consequently the full half-cycle of the power voltage cannot be applied to the load when the control voltage dictates that it should. This effect will naturally become most prominent at the maximum output end of the transfer characteristic curve, and it manifests itself as a reduction in maximum load current available, as shown in Fig. 5.

Rectifier leakage in the load circuit during the reset half-cycle will allow a current to flow through the gating winding in such a direction as partially to reset the core, giving the same detrimental results at full output as low remanence. Again this result becomes most evident at the full output end of the curve since at low outputs the voltage in the control circuit is transformed into the load circuit in such a direction as to reduce the back voltage across the load rectifier and in turn reduce the back leakage. This result is also shown in Fig. 5. Of course, if the maximum leakage current is not as large as the magnetizing current requirements in a square-loop material, then this reset theoretically will not occur, according to this analysis.

However, during dynamic operation there are other reverse current phenomena in rectifiers that occur in addition to the rectifier leakage as measured using reverse d-c voltage. Conrath of Vickers, Inc., has suggested that there is a capacitive component of reverse current. In tests on Vickers selenium rectifiers, he established at 60 cycles an equivalent capacitance value of 0.02 microfarad per square centimeter of rectifying area. Observations at this laboratory suggest that these equivalent capacitance values may vary with manufacturers of selenium rectifiers.

Since leakage and capacity effects should increase with rectifier plate area and frequency (as qualitatively checked in experiments) it becomes evident that circuits which operate satisfactorily for small maximum load currents at 60 cycles may not give designed full output when operated at 400 cycles with large current capacity rectifiers.

Circuit Theory

As a specific remedy for some of the ills encountered in the Ramey high-speed response magnetic amplifier when using square-loop materials and to extend the benefits of that circuit to amplifiers using the nonsquare-loop material, the circuit shown in basic and general half-wave form in Fig. 6 was devised.

The notation in Fig. 6 is as follows:

The polarities indicated for e_{ac}, e_{ac}', e_{ac}'' (the alternating supply voltages) are in-

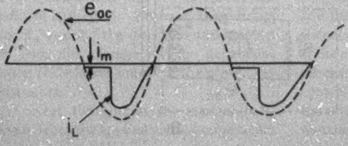
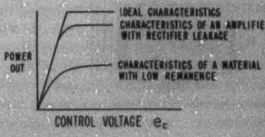


Fig. 4 (left). Current in load of half-wave magnetic amplifier

Fig. 5 (right). Transfer characteristics of half-cycle response magnetic amplifier using rectified alternating voltage control



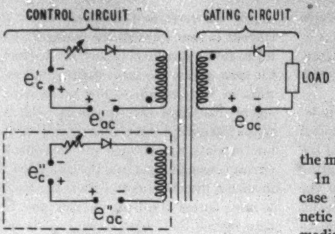
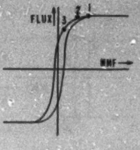


Fig. 6 (left). Half-cycle response magnetic amplifier using both resetting and presetting control circuits

Fig. 7. Hysteresis curve for intermediate remanence magnetic material



stantaneous polarities. These voltages possess the same relation to each other as the turns on their respective windings.

The control voltages, e_c', and e_c'', may be direct current, rectified alternating current, resistance drop, or any other proposed method of amplifier control.

The polarities of the windings are as indicated.

It is apparent that this circuit, with the exception of the element contained within the dashed enclosure, is the circuit shown in Fig. 1. The power delivered to the load on the gating half-cycle is determined by the amount of reset dictated by the control circuit on the reset half-cycle.

It is also evident that the element within the dashed enclosure is similar to the Ramey control circuit in applied voltage polarity and rectifier polarity, so that it may assume functions of a control circuit. It differs, however, in that its winding polarity is opposite to that of the Ramey control circuit, i.e., current through one control circuit changes flux level in the core in a direction opposite to that produced by current flow in the other circuit.

It will become evident later that in most circuits there must be a resistance in series with both control sources. The exceptions are those cases where a fixed minimum or maximum output might be acceptable or where the ratios of remanent to maximum flux is either 1.0 or 0.0. The values of the control source resistance will be influenced by the type of core, the rectifiers, and the type of control used. The circuit is drawn as shown to present the most general case.

In discussing the operation, a specific case using a core which possesses magnetic hysteresis characteristics intermediate between square-loop material and material possessing no remanent magnetism (see Fig. 7) will be assumed. Reset will be resistance controlled (Fig. 8A). The polarities of the supply voltages e_{ac}, e_{ac}' and e_{ac}'' are shown as they will be during the gating half-cycle. In the discussion, the Ramey control circuit will be called the reset circuit and the control circuit added will be called the presetting circuit.

In the step-by-step analysis of what occurs in this circuit, it might assist the understanding of the operation if it is first emphasized that the action may be divided into five distinct stages of operation. The first stage will be postulated as the firing stage, and the next two stages merely cover the short transition period into the fourth stage where the important circuit reactions occur. For understanding the circuit, stages two and three could be ignored if certain minor anomalies in state between stages 1 and 4 were accepted. Stage 5 covers the prefiring period in the gating half-cycle. For continuity the different modes will be discussed in their order of occurrence rather than their order of importance.

In stage 1 the core has saturated and current is limited only by the load impedance. No change of the state of the core is involved. As the instantaneous voltage decreases in the gating half-cycle, the current in the load circuit decreases until it reaches a value equal to the saturation magnetizing current, point 1, Fig. 7.

Stage 2 then begins. As instantaneous a-c voltage continues to decrease there is a change of flux in the core which generates voltages e_t, e_t' and e_t'' at the terminals of the windings on the core. These voltages will have positive polarities at the undotted ends of the windings. Thus the generated voltages will be directed against the rectifier in the resetting circuit and with the rectifier in the presetting circuit. However, as long as $|e_t''|<|e_{ac}''|$ the net voltage in the presetting circuit will be against the rectifier. Since the net voltages in both control circuits are against their respective rectifiers, the only active element is the load circuit. The voltage equation for the active element during stage two will be

$$e_t+e_{a-c}-i_LR_L=0 \tag{1}$$

During this stage the core will be partly demagnetized. When $e_t''=e_{ac}''$ the rectifier in the presetting circuit will unblock and stage three begins.

During the third stage current flows in the presetting circuit and the load circuit as determined by the equations

$$e_t''-e_{ac}''-i_c''R_c''=0 \tag{2}$$

$$e_t+e_{ac}-i_LR_L=0 \tag{3}$$

$$Ni_L+N''i_c''=Ni_m \tag{4}$$

where N, N', and N'' are the turns in the power winding, reset winding and preset winding respectively, and i_m is the total magnetizing current referred to the load circuit coil. As this stage progresses, the core will continue to demagnetize with the magnetizing current shifting to the presetting circuit until the supply voltages have reversed polarity and e_{ac} rise to a value equal to e_t. By this time, the magnetizing current has transferred to the presetting circuit, so current ceases to flow in the load circuit and that circuit becomes inactive. Stage 3 now ends and

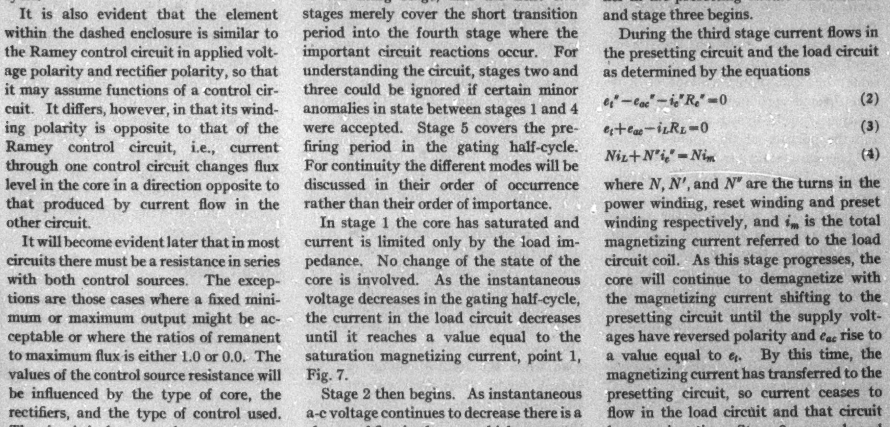


Fig. 8. Resistance controlled half-cycle-response magnetic amplifier

A (left). During gating half-cycle

B (right). During control half-cycle

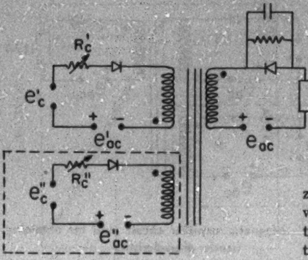


Fig. 9 (left). Magnetic amplifier with leaking power rectifier

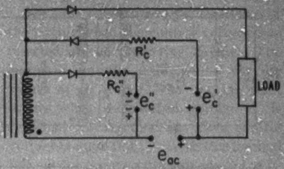


Fig. 10 (right). Single-winding half-wave magnetic amplifier incorporating both preset and reset circuit functions

the true resetting-presetting period will begin. At this time the core will be in a condition indicated by point 2, Fig. 7.

During stage 4, the polarities of the applied voltages will be as shown in Fig. 8(B). At the beginning of stage 4, $e_t' = e_{ac}'$, and e_{ac}' will continue to increase in absolute value faster than e_t' so the rectifier in the resetting circuit unblocks and starts to take part in the control of the core. Current is now flowing in both control circuits and the circuit equations become

$$e_{ac}' - e_t' - i_c'R_c' = 0 \tag{5}$$

$$e_{ac}'' + e_t'' - i_c''R_c'' = 0 \tag{6}$$

$$N''i_c'' - N'i_c' = Ni_m \tag{7}$$

From equations 5, 6, and 7, if 1:1:1 turns ratio exists and $e_{ac} = e_{ac}' = e_{ac}''$

$$e_t = e_{a-c}\frac{R_c'' - R_c'}{R_c'' + R_c'} + i_m\frac{R_c''R_c'}{R_c'' + R_c'} \tag{8}$$

It is in this stage that the state of the core at the end of the control cycle is determined by the relative values of the resistances. If the presetting circuit is open (R_c'' is very large), $e_t = e_{ac} + i_mR_c'$, and the resetting circuit will gain control to reset the core. This implies that e_t will not change sign during the control cycle. If the resetting circuit is open, $e_t = -e_{ac} + i_m R_c''$; and e_t will go through zero and change sign, as e_{ac} increases. Or, the core will go towards saturation. At the time when e_t goes through zero, the core state will have regressed down the hysteresis curve to point 3, Fig. 7, after which time the core will proceed towards saturation to point 1, Fig. 7. For least average current flow in the control circuit at full output, R_c'' should be chosen so that point 1, Fig. 7 is reached just at the end of the control half-cycle. Between these limits for R_c' and R_c'' the direction and amount of core flux change will be dependent upon the relative values of the control resistances.

Near the end of the control half-cycle, when the applied voltage approaches zero, a transition period will be entered which can be analyzed in a manner similar to that covered in stages 2 and 3 mentioned previously.

During this transition interval, magnetizing current will be gradually transferred to the load circuit. However, this period is so short, and the net applied voltage integral is so small, that there will be very little change of core state. At the end of this transition period the load circuit voltage will be acting upon the core, and the flux will start toward positive saturation from a point on the hysteresis curve essentially established during stage 4. When the integral of the voltage applied during the fifth stage is large enough, the core will saturate and drop out of the circuit. This returns the core to stage 1. For the more general case of resistance and/or voltage control, as shown in Fig. 6, the circuit equations developed in stage 4 will become

$$e_{ac}' - e_c' - e_t' - i_c'R_c' = 0 \tag{9}$$

$$e_{ac}'' - e_c'' + e_t'' - i_c''R_c'' = 0 \tag{10}$$

$$N''i_c'' - N'i_c' = Ni_m \tag{11}$$

So with 1:1:1 turns ratio and $e_{a-c} = e_{ac}' = e_{ac}''$

$$e_t = e_{ac}\frac{R_c'' - R_c'}{R_c'' + R_c'} + i_m\frac{R_c''R_c'}{R_c'' + R_c'} + \frac{e_c''R_c'}{R_c'' + R_c'} - \frac{e_c'R_c''}{R_c'' + R_c'} \tag{12}$$

This equation is similar in form to that previously given, except that there are now four parameters which will influence the final state of the core at the end of the control half-cycle. This is now a magnetic amplifier whose condition at the beginning of the gating half-cycle has been completely predetermined in the previous half-cycle of applied voltage, and this control may take place independently of the core material used.

If the material is perfectly square-loop, the preset circuit may be open-circuited, and the reset will be controlled by the resistance (or voltage) in the reset circuit. If the material has zero remanence, the reset circuit may be omitted and the preset controlled by the resistance or voltage in the preset circuit. A material which has a remanence below maximum flux level and above zero flux level could use either or both control circuits depending upon the desired range of output and the core characteristics.

If both circuits are used, the presetting circuit can be adjusted to give maximum output power with no signal furnished, and then signal voltage in the resetting circuit can be supplied to change the output power. Or the resetting circuit can be adjusted to give minimum output with no signal, and the signal voltage in the presetting circuit will increase output power. If desired, control can be inserted in both circuits. Control of the amplifier may be by any of the commonly used methods with the restriction that under most circumstances the signal source must have either an external or internal series resistance.

Another source of benefits arising from the use of the circuit presented here is the compensation for rectifier back leakage when using square-loop material.

In many circuits with small magnetizing current requirements the back leakage through the power rectifier causes uncontrolled reset when no reset is desired. This condition is aggravated when larger power outputs are desired, as previously discussed. This uncontrolled reset is quite similar to the effect observed when the core material has a remanent flux less than maximum flux.

Fig. 9 shows the magnetic amplifier during the reset half-cycle. The d-c back leakage of the rectifier is portrayed as the resistor and the a-c dynamic leakage is portrayed as the capacitor in parallel with the power rectifier.

If it is assumed that the material is square-loop, and full output is desired, the first requirement is that the signal voltage, e_c', be equal and opposite to applied e_{a-c}' (if voltage control of reset is used).

Indicated polarities of the windings and applied a-c show that, if back leakage current of a greater value than the magnetizing current tries to flow in the load circuit, reset will occur which will not be under control of the reset winding.

But in the presetting circuit the generated voltage, e_t'', will be with the rectifier, and current can flow in the circuit, minimizing the amount of reset. In addition,

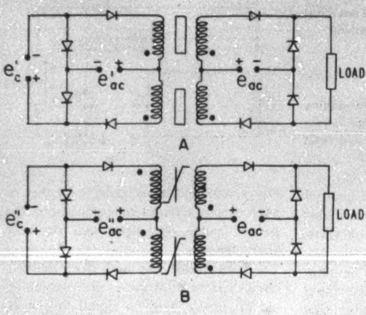


Fig. 11 (left). Full-wave magnetic amplifier

A. Using square-loop magnetic materials

B. Using zero-remanence magnetic materials

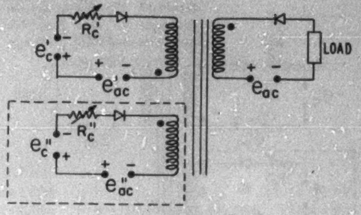


Fig. 13. Magnetic amplifier circuit used for obtaining transfer characteristics

the applied voltage, e_c'', can give a positive current flow in this winding in such a direction as to resist demagnetization of the core. If some reset of the cores does occur during the first part of the reset cycle from the capacitive component pulse of rectifier leakage, the current flow in the preset circuit, as forced by e_{ac}'', can reset the core to the saturation point and cause the core to be fully saturated when the power cycle begins. The full half-wave of voltage in the gating half-cycle will thus be made available to the load.

Although, to add clarity to the discussion, the circuit heretofore has been presented as though three separate windings were placed on the core, the circuit will operate in the same way if only one winding is used and the control circuits added to that winding as shown in Fig. 10. The circuit in Fig. 10 is not strictly analogous to that presented in Fig. 6, since e_{ac} is not placed in the presetting circuit, but its functions are taken over by e_c''. The relative polarity of the signal voltage in the presetting circuit, e_c'', appears ambiguous because it will depend upon core material and desired circuit operation; i.e., if the material is one with zero remanence, the resetting circuit would be omitted and the presetting polarity would be in such a direction as to buck the voltage generated by the flux collapse. If the material is square-loop with rectifier leakage, the polarity of the voltage e_c'' would be reversed to provide a positive presetting of the core to the saturated condition.

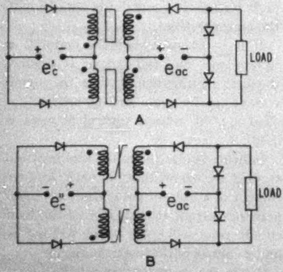


Fig. 12. Full-wave magnetic amplifier circuit

A. Using square-loop magnetic materials
B. Using zero remanence magnetic materials

The resetting-presetting circuits presented so far have been in the half-wave form. They may also be used in full-wave form with the stipulation that some types of control may require changes in the circuit configuration because of the possibility of disturbing voltages from one element acting upon the other element. It is emphasized that among other factors the approach to design of full wave circuits depends upon whether the material has square-loop or zero remanence characteristics. For example, the circuit given in Fig. 11(A) for square-loop material with d-c control appears similar to that given in Fig. 11(B) for zero remanence material, but closer examination shows different functions for the various voltages.

In Fig. 11(A), e_{ac}' will reset the core in absence of control voltage, e_c', giving zero output. However, in Fig. 11(B), e_{ac}'' has only the purpose of supplying a slight boost voltage to overcome IR losses in the presetting circuit. If zero resistance in the control circuit is assumed, e_{ac}'' should be zero.

In the absence of voltages e_c'' and e_{ac}'', in the control circuit, the zero remanence core in Fig. 11(B), which is on the presetting cycle, will generate a voltage causing the magnetizing current to circulate through the control winding. The core is now acting as an inductance with infinite time constant, and refusing to reset from the saturated condition. Maximum output current would be available to the load on the gating cycle. As the control voltage, e_c'', increases, it will force the core to reset an amount equal to the voltage integral of the signal, until at full signal the output current would be equal to the magnetizing current.

A more dramatic illustration of the required difference in approach is shown in Figs. 12(A) and 12(B) for full-wave operation with d-c control. Although the circuits appear similar, examination shows that Fig. 12(A) will not operate because both cores are being acted upon simultaneously; and with element polarities as shown, there would be no alternation of reset between the two. However, with the circuit in Fig. 12(B), the flux collapse in one core during the time that core is being preset generates a voltage equal e_c'' (which voltage is blocked by the rectifier in the other core circuit). Thus full wave operation may be obtained with the zero remanence material with the circuit in Fig. 12(B), whereas a different circuit must be used for the square-loop material.

Reduction to Practice

As a check of the operation of this circuit, two toroidal cores of different hysteresis characteristics were wound and successively placed in the circuit shown in Fig. 13. Transfer characteristic curves were then obtained for these half-wave amplifiers.

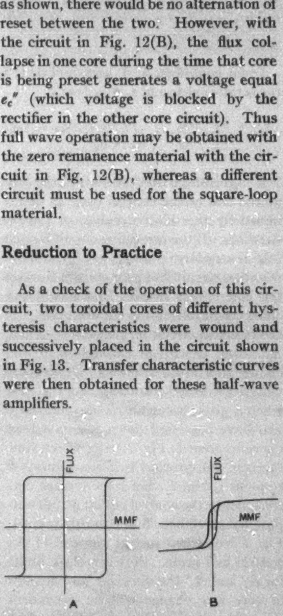


Fig. 14. Hysteresis curve

A. For core no. 1
B. For core no. 2

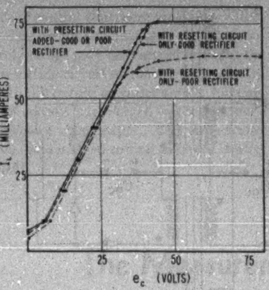


Fig. 15. Transfer characteristics for core no 1

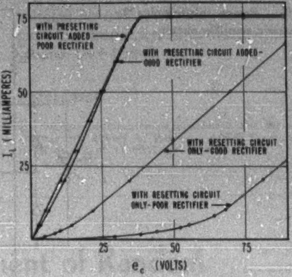


Fig. 16. Transfer characteristics for core no. 2

Core no. 1 was a grain-oriented silicon-iron material which has a relatively broad square-loop, with a published coercive force of 0.2 to 0.4 oersteds and a maximum flux density of 14,000–15,000 gauss. Core no. 2 was an alloy of 5 per cent molybdenum and 79 per cent nickel, with a relatively narrow hysteresis loop of 0.02 to 0.04 oersteds, a maximum flux density of approximately 6,600 gauss and a remanent flux density of approximately 5,000 gauss. The relative size and shape of the hysteresis loops are shown in Figs. 14(A) and 14(B).

Rectifiers were of a selenium 4-plate half-wave type, rated at 0.25 ampere. However, during the tests involving the effects of poor rectifiers a selenium rectifier of the same rating from a different manufacturer was placed in the load circuit. In this discussion "poor rectifier" means only that back leakage effects are greater, with deleterious effects on the action of the magnetic amplifier.

Each core was wound so as to saturate at 40 volts, 60 cycles alternating current. The circuits were arranged as half-wave amplifiers with rectified alternating voltage control, since this type of control gives a straight-line transfer characteristic curve when the amplifier is operating properly. Thus any deviations from the ideal curve may be easily noticed. This type of control was used in reference 3.

Tests were first run using the resetting circuit only (presetting winding open). Then the presetting circuit was closed with a resistance in series of a proper value to give full output when the net resetting voltage in the resetting circuit was zero, and the test runs repeated. The next step was to take readings using the resetting circuit only, after inserting a rectifier which was not as suitable for the circuit. This run was then repeated after the presetting winding was placed in the circuits.

Fig. 15 shows the results obtained while using core no. 1. It may be observed that the original transfer characteristic for this core approximated quite closely the ideal curve as shown in Fig. 5. However, when the poor rectifier was substituted in the power circuit, the maximum output decreased considerably and a curve was introduced at the knee. When the presetting circuit was added to the core, the curve returned to its former shape and the maximum output increased to its previous value.

Figure 16 shows the results obtained while using core no. 2. It is seen that while using the resetting circuit only the transfer characteristic does not approach the ideal curve even with good rectifiers, and it is also seen that the curve is seriously affected by the type of rectifier. When the presetting winding is added, the transfer characteristic curve again duplicates the transfer characteristic obtained using a square-loop material and good rectifiers, exclusive of the rectifier type. This is also shown in Fig. 16.

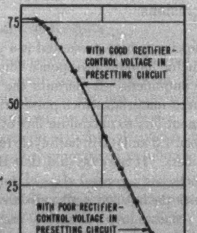


Fig. 17. Transfer characteristics of core no. 2 with control voltage in presetting circuit

The next concern was the operation of the circuit when the control voltage was inserted in the presetting circuit. The circuit is initially adjusted by varying values of R_c' and R_c'' so that no reset occurs when the control voltage e_c'' is zero and full reset occurs when control voltage is equal to e_{ac}''. With this method full output occurs with zero control voltage, as is shown by transfer characteristic curves in Fig. 17. Again it may be noted that little difference is evident in the curve regardless of whether the poor or the good rectifier is used in the power circuit.

As another example of the versatility of this circuit, it was decided to insert the control voltage in the presetting circuit so as to give zero output with zero control voltage. e_{ac}'' was omitted in the presetting circuit and the reset circuit resistance value adjusted to give full reset with zero control voltage. Then, as the control voltage was increased, the output increased (see Fig. 18). These curves have the same general slope as those in Fig. 17 with reversed sign.

As may not be surprising, these improvements have not been realized without some penalties. In this case it is evident, both from the equations and from intuition, that the input current on the reset cycle must be increased since from equation 7 $N''i''_c - N'i'_c = Ni_m$ and the relative values of i_c' and i_c'' will be determined by R_c' and R_c'' whose values in turn will depend upon core material and rectifiers used. Practically, however, the increased input power requirement is not detrimental in many cases. This is especially true where resistance control is used. If R_c' and R_c'' are caused to vary individually, the circuit requirements are such that either the reset or preset circuit

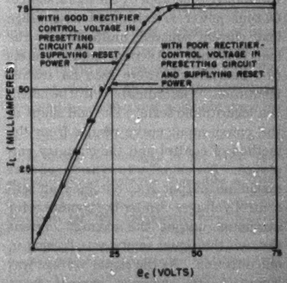


Fig. 18. Transfer characteristics of core no. 2 with control voltage in presetting circuit and supplying power

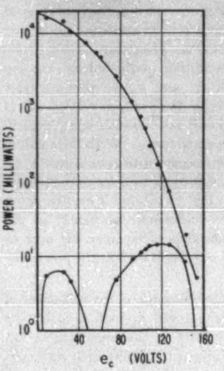


Fig. 19 (left). Power relationships in magnetic amplifier circuit

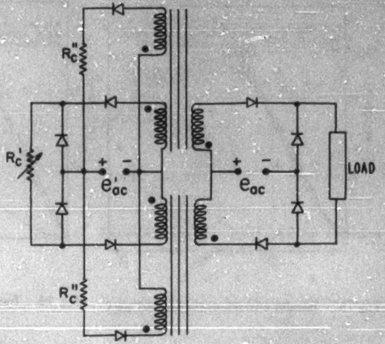


Fig. 20 (right). Full-wave magnetic amplifier circuit for exciter

will be effectively open when the other element is controlling the amplifier. Consequently the input power requirements have not been increased. However, if it is desired to leave one resistance fixed, and to vary the resistance in the other element, power requirement increases will still be well within practical limits. For outputs of many watts the input current can still be only a few milliamperes.

Rather than discuss the amplifier in terms of power gains (which can become a difficult subject when examining the half-cycle response circuit) the curves in Fig. 19 present the input power and output power for increasing signal voltage as obtained on a 5-79-molybdenum nickel alloy core with resistance load and rectified alternating voltage control in a half-wave circuit. It is realized that there are many ways of presenting power gain in amplifiers, but in this discussion input power is presented as the average current through the signal source, multiplied by the voltage generated by the signal source. Output power is calculated as average load current squared, multiplied by the load resistance. Average reading meters (d-c instruments) were used for instrumentation. The ordinate, it will be noted, is a logarithmic scale. The odd shape of the power input curve results from the method of control and the circuitry employed. The circuit was so arranged that maximum output was obtained at zero control voltage. Under the condition for maximum output the average current through the signal source was flowing in one direction. As the signal voltage was increased, the average current went through zero and reversed direction. It then increased in the opposite direction for a while before it started decreasing again to zero. No claim is made that circuit elements used were the optimum for high power gain. The results are merely presented as an example.

The circuit presented herein has proved its practicability when incorporated in a resistance controlled exciter using magnetic amplifiers with nominally square-loop materials. The circuit used is shown in Fig. 20. Originally maximum output current in the power circuit could not be obtained because uncontrolled reset occurred before addition of the presetting circuit. Maximum available current in the load circuit was increased by 70 per cent with its addition. In furtherance of the previous contention that, although input current requirements are increased, the current will remain within practical limits, it was noted in this case that when the available output current was increased to 3.4 amperes maximum, the maximum current in the control circuit was still less than 9 milliamperes at zero output. This condition entailed maximum input current requirements.

Conclusions

The circuit here presented is a general circuit for half-cycle response time magnetic amplifiers. It permits the use of magnetic materials which possess ratios of remanent flux to maximum flux of values between 1.0 and 0.0 in half-cycle response amplifiers. It may be said that this circuit presents a general class of amplifiers of which the half-cycle response amplifiers using square-loop material previously reported upon represent a special case. Another special case would be the use of magnetic materials with zero remanence in half-cycle response amplifiers. This circuit encompasses both cases, and provides for the use of materials possessing intermediate characteristics.

In addition, compensation for some defects encountered in the special case amplifiers using practical circuit elements may be achieved with this circuit. Among these defects are those caused by some loss in remanent magnetism in square-loop materials, or some residual magnetism in saturable inductance materials. Related to these effects and similar in nature are those caused by rectifier back leakage. These two effects may be simultaneously compensated for in this amplifier.

In general, the input power is an inverse function of the maximum resistances which may be placed in the control circuits. As a consequence, the input power requirements will decrease as the remanent flux level approaches either zero or maximum flux level, all other factors remaining constant.

The foregoing factor leads to the conclusion that if narrow hysteresis loop materials could be manufactured with zero remanent flux level, the half-cycle response circuits could be designed with a much lower input power requirement than the present square-loop materials. Preliminary work is being conducted on a system which will use magnetic materials with low coercive force requirements such as are commercially available. The zero remanence condition which is desirable for low input power will be simulated by introducing a constant current bias.

However, it must be realized that this circuit is not an absolute panacea. Although its use will enable full power output in amplifiers employing non-square-loop magnetic materials, and leaky power circuit rectifiers, the input power requirements may be raised.

The circuit responds best as far as power gain is concerned when resistance control is used but it is perfectly feasible to use voltage control or any of the methods of control that were previously cited.

The circuit is versatile in that the control voltage may be inserted in either the presetting circuit or the resetting circuit, with full output at either zero or full control voltage, as arranged by the designer's choice of components. The circuit may even be arranged so that different control voltages are inserted in the two circuits providing the restrictions as to source impedance are followed.

References

1. ON THE MECHANICS OF MAGNETIC AMPLIFIER OPERATION, Robert A. Ramey. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1214–23.

2. ON THE CONTROL OF MAGNETIC AMPLIFIERS, R. A. Ramey. *Ibid.*, pp. 2124–28.

3. THE EFFECT OF CORE MATERIALS ON MAGNETIC AMPLIFIER CIRCUITS, Leo J. Johnson. *AIEE Transactions*, vol. 71, pt. I, Jan. 1952, pp. 28–31.

◆

No Discussion

The Measurement of Random Monochrome Video Interference

J. M. BARSTOW
MEMBER AIEE

H. N. CHRISTOPHER
NONMEMBER AIEE

ONE OF the important factors in the design and maintenance of television transmission circuits is random interference, sometimes referred to as random noise. Much of this type of interference originates in the input stages of amplifiers but it may also originate in the line or in other parts of the transmission path. These noise components are greatly modified by circuit equalization which may vary from a few decibels (db) to possibly 27 or more db per octave. Thus the transmission engineer is confronted with the problem of evaluating random interference having different energy levels, per cycle, over the television bandwidth. Before noise evaluations can be made, information must be obtained by subjective tests on the interfering effect of broad, narrow, and mixed bands of random noise distributed throughout the television band.

Out of the work associated with such tests comes a reasonable question, "Cannot random interference be measured in such a way that equal measurements will mean approximately equal interfering effect, regardless of the frequency composition of the noise?"

This question may be reasonably divided into two parts:

1. What is the weighting or relative importance of random interference in different parts of the video spectrum?

2. Does the human visual mechanism sum up the interference effects of random noise in various parts of the video spectrum in such a way that the over-all effect can be uniquely related to an over-all measurement, made with equipment that is not too complicated?

Preliminary answers to these questions are given in the following sections of this paper. The answers are called preliminary because it is obvious that they depend to a large extent on the equipment used in making the judgment tests, and this equipment is undergoing constant change as the television art advances. At present, it is believed that a stage has been reached at which results of the type involved will be useful over a sufficiently long period to make publication worth while.

Others have considered this problem and have deduced approximate results by rationalizing a relatively small number of subjective measurements with related work involving photographic reproduction processes.[1,2] The test results given here, however, are purely experimental and seem sufficient in scope and numbers to warrant conclusions. Although they confirm, in part, the general characteristics of random noise weighting previously estimated, they are independent of any thus far published.

Summary of Results

The frequency weighting derived from the judgment tests is given in Fig. 1. This weighting is for use with a simple power-summing measuring device.

The general principles derived from the tests may be summarized as follows:

1. Low-frequency noise is judged much more interfering than high-frequency noise of equal power.

2. A given amount of noise power is judged more objectionable if it is concentrated in a narrow band than if it is spread out over a wider band in the same frequency region.

3. Human vision in combination with the present television monitors does not precisely sum weighted noise powers in arriving at an over-all assessment of the interfering effect of random noise bands. However, a reasonable compromise can be obtained with weighting applied to a power meter. In the region of 7 db above threshold, average errors of the order of 1 db with maximum errors of 2 db will obtain.

4. At frequencies above 4.5 megacycles (mc) an unexpected effect was observed called sparkle effect. When the total noise power is contained in the region above 4 mc, sharp points of light of very brief duration appear on the raster. It is believed that this effect is the result of the random occurrence of high peak potentials in the random noise which produce the sharp points of light on the raster before the extremely fine-grain noise effect becomes visible. The sparkle effect tends to flatten the weighting above 4 mc.

Subjective Test Conditions

The objective in setting up the test conditions was to simulate somewhat more severe viewing conditions than are usually encountered in the home. The test results were obtained by a jury of observers viewing the raster from a distance of four times the picture height. The raster was adjusted to 6 by 8 inches and was viewed in almost total darkness (ambient about 0.0025-foot lambert). The jurors were subjected one at a time to the test conditions, and for most tests three juries of 10 observers each were employed. The three juries had members in common, there being 14 different jurors used in all.

The test setup is illustrated schematically in Fig. 2. A test consisted of having the juror compare the judgment criterion (7.22 mc low-pass band of flat noise at about 7 db above the threshold of visibility) with a test noise obtained by

---

Paper 53-356, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Pacific General Meeting, Vancouver, B.C., Canada, September 1–4, 1953. Manuscript submitted June 3, 1953; made available for printing July 13, 1953.

J. M. BARSTOW and H. N. CHRISTOPHER are with the Bell Telephone Laboratories, Inc., New York, N. Y.

The authors wish to acknowledge the very able assistance given by A. D. Fowler of the Bell Telephone Laboratories, Inc., in reviewing the test results and providing valuable comments during the preparation of the material for this paper.

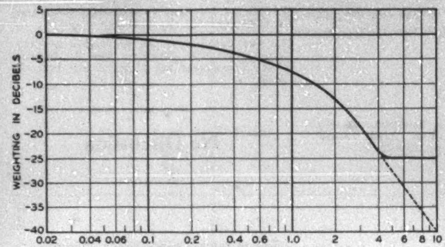
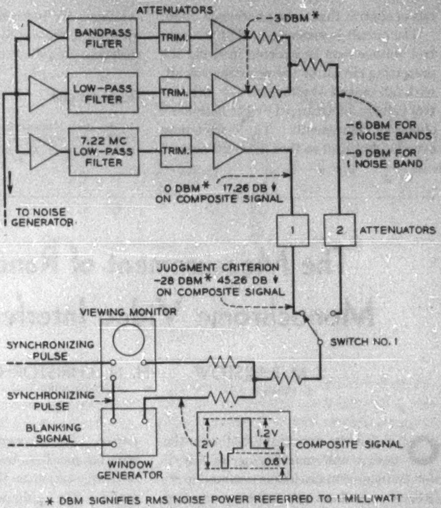


Fig. 1. Random noise weighting for monochrome television



* DBM SIGNIFIES RMS NOISE POWER REFERRED TO 1 MILLIWATT

Fig. 2 (right). Circuit used to determine the relative interfering effect of discrete bands of random noise

passing a wide band of flat noise through either a bandpass or a low-pass filter. The test level of about 7 db above the threshold of visibility (45.26 db below peak-to-peak composite signal) was chosen for the reference condition because it is in the region of a tolerable noise level, and therefore information on methods of measurement which apply to it would be most useful. (Zero db obtain when the ratio of rms noise volts to composite signal voltage, peak to peak, =1). The observer was permitted to switch (switch 1, Fig. 2) from one condition to the other at will, and was instructed to ask the test operator to raise or lower the level of the test noise as required to obtain interference equality in his judgment. The circuit arrangement of Fig. 2 also permitted the mixing of two noise bands in different parts of the video spectrum for test.

In the early stages of noise tests, picture slides were used as subject material. It was soon observed, however, that the jurors were basing their judgments on the noise in middle gray areas of the picture and check tests were made on wholly gray areas to corroborate this observation. Subsequently the tests were made on wholly gray areas without the picture. This technique permitted the raster to be generated electronically without the use of a scanner, and thus provided essentially a noise-free raster, a feature greatly desired in noise studies.

It was found that about equal results were obtained over a range of gray areas varying in brightness from about 0.3-foot- to 0.6-foot lambert.

The types of random noise tested varied from very wide-band noise (7.22-mc reference condition) to narrow bands only 40 kc wide. Although no band used was narrow enough to produce definite line patterns on the raster, some bands were sufficiently narrow to produce fleeting appearances of herring bone patterns.

The tests on noises having two bands in different parts of the spectrum were made, to obtain additional evidence on the rule of combination of interfering effects as summed up by human vision. Noises typical of those encountered on standard transmission facilities were also included in the tests.

All viewing was on a laboratory monitor equipped with an *1816 P4* kinescope. The kinescope anode was operated at 11 kv and the associated video amplifier was substantially distortionless to above 15 mc. The simulated picture signals were in accordance with the present United States standard of 525 lines, 60 fields per second, 30 frames per second, with alternate field interlace.

The electrical transmission circuits were uniform in their response up to about 8 mc. This means that the monitor was superior to present-day broadcast television receivers in its response in the 0-4.3-mc range, and very much more uniform in response in the range of 4.3-8 mc. The monitor and associated circuits were made good in this high range to permit extension of the information on relative interfering effect toward 8 mc. Hence, it is believed the weighting and measuring methods to be described apply to the present standard monochrome system up to 4.3 mc, and give some indication of the weighting which should be applied to interference on wider circuits used for kinescope recording and wide-band theatre television employing 525-line 15,750-lines-per-second standards.

Discussion of Results

The results of the tests are shown in Figs. 1, 3, 4, 5 and 6, and Tables I to III.

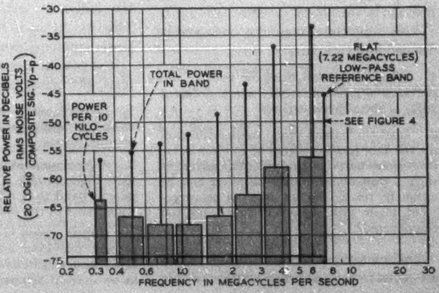


Fig. 3. Powers in bands of noise judged to be equally interfering (30-observer data)

Table I. Unweighted Noise Powers Judged Equally Interfering*

30-Observer Bandpass Data of May 20, 1952, and July 18, 1950

Passband, Megacycles (Mc) Mid	Passband, Megacycles (Mc) Full	Total Power, Decibels (Db)	Power/10 Kc, Db
0.325	0.050	−56.8	−63.8
0.505	0.165	−55.4	−66.6
0.750	0.264	−53.9	−68.1
1.120	0.388	−52.3	−68.1
1.710	0.600	−48.8	−66.6
2.55	0.889	−43.5	−62.9
3.80	1.29	−36.9	−58.0
6.15	2.14	−33.4	−56.4

	30-Observer Low-Pass Data of May 20, 1952 and July 18, 1950		9-Observer Low-Pass Data of July 18–27, 1952	
Upper Cutoff, Mc	Total Power, Db	Power/10 Kc, Db	Total Power, Db	Power/10 Kc, Db
0.040	−59.8	−65.8		
0.080	−59.3	−68.4	−58.8	−67.8
0.133	−58.4	−69.6	−57.7	−68.9
0.211	−57.9	−71.1	−57.5	−70.7
0.320	−56.5	−71.5	−56.6	−71.7
0.473			−56.1	−72.3
0.501			−56.4	−73.4
0.585	−54.9	−72.6		
0.734			−54.7	−73.4
0.890	−53.4	−72.9		
1.215			−53.0	−73.8
1.360	−52.1	−73.4		
1.970			−50.9	−73.8
2.300	−49.8	−73.4		
3.075	−48.6	−73.4		
4.73	−46.6	−73.4	−47.1	−73.8
7.22	−45.0	−73.6	−45.3	−73.8
Judgement Criterion	−45.3	−73.8	−45.3	−73.8

* The indicated powers were judged to have the same interfering effect as the 7.22-mc flat noise band used as a judgment criterion $20 \log_{10} \frac{\text{Rms noise volts}}{\text{Peak-to-peak composite signal volts}} = -45.26$ db

Fig. 3 shows the magnitudes of unweighted noise powers in the bands of noise which were judged to be equally interfering. The bandwidths indicated were selected to be narrow with respect to the full band but not narrow enough to cause bar pattern effects. The total power in each band is shown plotted near the center of each band. The point representing the total power in each band is the integral of the power per kilocycle over the area defined by the frequency boundaries of each band. For simplicity and ease of computation, the filter characteristics have been idealized. This figure shows that more power is required at high than at low frequencies to produce a given interfering effect. The inverse of the shape revealed by the total power points is very nearly equal to the shape of the weighting given in Fig. 1.

The unweighted power in the low-pass reference band is indicated in Fig. 3 by a point plotted at 7.22 mc, the upper edge of this band. This level is 45.26 db below the peak-to-peak composite signal. Because of the unsymmetrical nature of television video signals, it has become customary to refer to the signal in terms of the peak-to-peak magnitude (picture plus sync) while the noise is expressed in rms magnitudes.

Figs. 4 and 5 show the unweighted powers in low-pass noise bands which were judged to be equally interfering. These data were obtained approximately 1 year apart and, although two setups and slightly different noise bands were used, the data are in very good agreement. The total unweighted powers are plotted at the upper edge of the low-pass bands in these figures. Considering the rate of increase in the total noise power required to produce the reference interfering effect as the bandwidth is increased, these data show again that on a per-cycle basis, high-frequency noise power does not contribute as much to interfering effect as does low-frequency noise power. In these figures the inverse shape formed by the total power points does not approximate the best weighting shape. This may be better understood when it is pointed out that in an area where a positive slope of 3 db per octave is attained, the noise powers in that area would be contributing nothing to the interfering effect. This slope is very nearly attained in the 2-to-7-mc region.

The horizontal lines showing the noise power per 10 kc in each band may be used to yield information on weighting also. A negative slope of 3 db in power level indicated by these lines, per octave of total

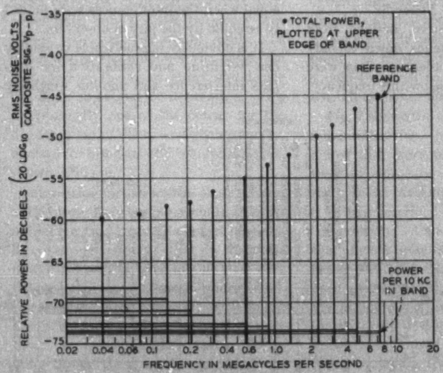


Fig. 4. Powers in low-pass bands of noise judged to be equally interfering (30-observer data)

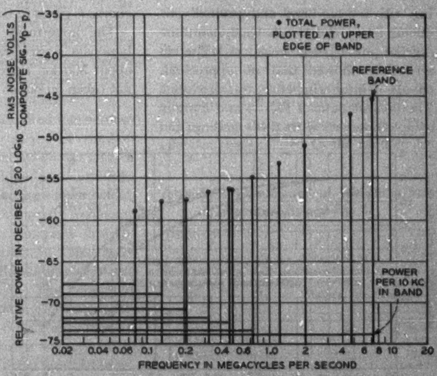


Fig. 5. Powers in low-pass bands of noise judged to be equally interfering (9 observer data)

Table II. Description of Noises Tested

Noise Test		Weighted Power, Db	Deviation from Average, Db	Deviation Squared, Db
1	Simulated *TD-2*, 12 db per octave maximum at 7.22 mc	57.05	−0.95	0.900
2	*L3* noise, down 10 db at 0.5 mc, 10 db at 1 mc, 7 db at 2 mc, 0 at 4.4 mc	56.45	−1.55	2.400
3	Flat noise, uniform energy per kc, 0 to 7.22 mc	56.43	−1.57	2.460
4	Two noise bands, see Table III	57.47	−0.53	0.281
5	Two noise bands, see Table III	57.86	−0.14	0.0196
6	Two noise bands, see Table III	57.76	−0.24	0.0576
7	Bandpass band centered at 0.325 mc full band 0.050 mc	60.00	+2.0	4.000
8	Bandpass band centered at 0.505 mc full band 0.165 mc	60.06	+2.06	4.250
9	Bandpass band centered at 0.750 mc full band 0.264 mc	60.02	+2.02	4.080
10	Bandpass band centered at 1.120 mc full band 0.388 mc	60.52	+2.5	6.250
11	Bandpass band centered at 1.710 mc full band 0.600 mc	60.24	+2.24	5.000
12	Bandpass band centered at 2.550 mc full band 0.889 mc	60.10	+2.1	4.410
13	Bandpass band centered at 3.800 mc full band 1.290 mc	59.28	+1.28	1.640
14	Bandpass band centered at 6.150 mc full band 2.140 mc	58.30	+0.30	0.090
15	Low-pass band 0 to 0.040 mc	59.80	+1.80	3.250
16	Low-pass band 0 to 0.080 mc	59.50	+1.50	2.250
17	Low-pass band 0 to 0.133 mc	58.80	+0.80	0.640
18	Low-pass band 0 to 0.211 mc	58.60	+0.60	0.360
19	Low-pass band 0 to 0.320 mc	58.12	+0.12	0.0144
20	Low pass band 0 to 0.473 mc	58.22	+0.22	0.441
21	Low-pass band 0 to 0.501 mc	58.61	+0.61	0.3720
22	Low-pass band 0 to 0.585 mc	58.35	+0.35	0.1220
23	Low-pass band 0 to 0.734 mc	57.65	−0.35	0.1220
24	Low-pass band 0 to 0.890 mc	56.90	−1.10	1.2100
25	Low-pass band 0 to 1.215 mc	57.12	−0.88	0.7750
26	Low-pass band 0 to 1.360 mc	56.50	−1.50	2.2500
27	Low-pass band 0 to 1.970 mc	56.72	−1.28	1.6400
28	Low-pass band 0 to 2.300 mc	56.15	−1.85	3.4260
29	Low-pass band 0 to 3.075 mc	56.05	−1.95	3.8100
30	Low-pass band 0 to 4.73 mc	56.15	−1.85	3.4300
31	Low-pass band 0 to 7.22 mc	56.17	−1.83	3.3500
		1801.00		62.8982
		58.00 average		2.03
				1.42 db standard deviation

bandwidth would indicate a frequency region in which the weighting is constant. The lower two or three low-pass bands approach such a region.

It is interesting to observe the powers per 10 kc in the low-pass bands having upper edges above 1 mc. These powers per 10 kc are nearly constant, indicating that, in the presence of the low-frequency noise power, the high frequencies are contributing little to the observed interfering effect. In some of the earlier work on the interfering effect of flat random noise, tests were made by starting with a 4-mc band and gradually reducing its width towards 1 mc, maintaining the power-per-kc constant. The observations showed that no appreciable change in interfering effect was produced. The present data of Figs. 4 and 5 are in complete accord with these findings but they also show a marked increase in the interfering effect of noise in the region below 1 mc. The early tests were not carried far enough to divulge this information.

Table I shows the magnitudes of the unweighted noise powers plotted in Figs. 3, 4, and 5.

Fig. 6 shows three weighting curves, one derived from the bandpass data, another derived from the low-pass data and a third or compromise curve drawn in between the other two. An explanation of the derivation of the top and bottom curves is given in the Appendix.

The uppermost curve, derived from bandpass noises, extends down to only about 300 kc, since adequate data were not taken on bands below this frequency. The bottom curve, derived from low-pass data, is not shown beyond 1.3 mc because in the presence of low-frequency interference, the determination of the position of the high-frequency end of the curve is subject to large errors, and hence there is considerable uncertainty as to its proper level, based on the low-pass data alone.

The curves are so placed on the db scale that if the top curve is used for the bandpass noises and the bottom curve for the low-pass noises, computed weighted noise powers turn out to be approximately equal when the interferences were judged to be equal by the jury of observers.

It is obviously impracticable to employ one weighting and calibration for low-pass interference and another weighting

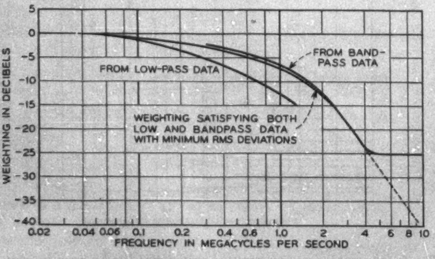


Fig. 6. Weighting shapes derived from judgment tests

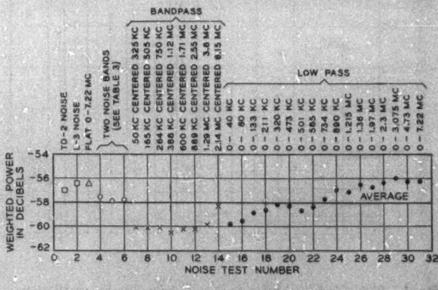


Fig. 7. Weighted powers of noises judged to be equally interfering

Table III. Results of 2-Band Noise Tests

Observations Made by 10 Observers

Noise Test No.*	Filter	Passband, Mc		Unweighted Power in Each Band, Db	Weighted Power in Each Band, Db	Total Weighted Power Combined Bands, Db
		Mid	Full			
4	Low-pass		0.133	−59.17	−59.77	
	Bandpass	1.12	0.388	−53.06	−61.26	−57.47
5	Low-pass		0.133	−60.06	−60.66	
	Bandpass	3.80	1.29	−38.56	−61.04	−57.86
6	Low-pass		0.211	−59.16	−60.16	
	Bandpass	2.55	0.889	−44.76	−61.36	−57.76
3 Judgment criterion†	Low-pass		7.22	−45.26	−56.43	−56.43

* Test numbers as listed in Table II.

† The power in the combined bands was judged to equal the interfering effect of the 7.22-mc low-pass band used as a reference or judgment criterion.

and calibration for bandpass interference. Hence, a compromise is necessary if the summation of weighted power is used as an indication of interfering effect.

At this point it should be noted that the different positions of the two curves reveal that for the interference levels tested, human vision does not sum the weighted interference powers over the band in assessing over-all interfering effect. The relative positions of the curves indicate that a given amount of noise power is more interfering if it is concentrated in a narrow band (bandpass data) than if it is spread out over a wider band (low-pass data) in the same frequency region.

Consideration of the interfering effects of narrow- and wide-band noises, having the same power, in the same general frequency region, indicates that the average observer saw something that caused him to rate the narrow bands more interfering. It is a fact that narrow bands tend to produce fleeting patterns not very unlike herringbone effects. If the band could be made progressively narrower until only a single frequency was passed, bar patterns would result. In other words, the narrower the band, the less random the noise effect and hence the more objectionable it becomes. This is apparently one of the causes for the displacement of the two weighting curves, derived for low-pass and bandpass noises, shown in Fig. 6.

Theorizing further on the observations, and noting that they were made at levels only a few db (about 7) above threshold, it appears reasonable that if the noise power in a narrow band at this level were redistributed over a much wider band, the resulting noise power per cycle might easily be so low as to be barely above threshold, or even slightly below threshold. Under these conditions a balance between the interfering effects would be obtained only by raising the power in the wide-band noise relative to that in the narrow-band noise. This reasoning offers a further explanation of the observed results.

The curve plotted between the two outside curves in Fig. 6 is a compromise weighting which produces nearly constant computed weighted powers when different shapes of random interference are adjusted to be equally interfering by competent jurors. The philosophy involved in drawing this curve between the other two may be described as follows: At low frequencies, where the only evidence available is from noise obtained through low-pass filters, the compromise curve is near the low-pass curve. At high frequencies where the low-pass data gives very poor evidence of what the weighting should be, the compromise curve is drawn near the bandpass curve. This is done not only because the best evidence on what the weighting should be at high frequencies is obtained from the bandpass data, but also because in actual practice most random noises are tipped up to some extent, and therefore approach, in the limit, bandpass noise having high noise power per cycle near the upper edge of the band. It would appear that full weight should be given to high-frequency noise when it occurs in this manner.

Fig. 7 shows the computed interference magnitudes for the series of random noises that were used in the course of these experiments. Each position along the abscissa presents a given noise, which is briefly described as regards frequency composition. A more complete description of each noise is given in Table 2, together with numerical magnitudes of the data plotted in Fig. 7. The computed magnitude using the compromise weighting is shown as the ordinate. Since all noises were judged to be equally interfering by jurors, the departure of the respective computed values from a constant average value represents the errors which might be expected from a meter designed to sum the weighted power according to the compromise weighting in Fig. 6.

An examination of Fig. 7 shows that narrow-band noises will in general be underweighted (the meter will read too low), and broad bands will be overweighted (the meter will read too high).

The rms deviation of the db magnitudes shown in Fig. 2 from the average value is 1.42 db. This may be compared to the rms deviation from the mean of all observer judgments on the various noises. The latter figure is 1.45 db. From this it may be seen that the meter error (assuming the mean judgment of the jurors to be correct in every case) is of about the same order of magnitude as the observer deviation.

Sparkle Effect

Fig. 1 shows two weightings for the area above 4 mc. The lower weighting is an extrapolation of the weighting determined by the judgment tests for frequencies below 4 mc, while the upper weighting more nearly represents the weighting determined by judgment tests on the noise band in the 4–8-mc region.

The appearance of noise as observed on a picture tube varies enormously, depending upon the frequency composition. Very low-frequency noise creates a streaky, nebulous effect that is judged most annoying. As the frequency increases, the disturbance resembles photographic grain in which the individual grains have random motion and become finer and finer as the frequency increases. As the grain becomes finer, the input noise level must be increased to keep the fine grain visible. In the vicinity of 4.5 mc an abrupt change occurs. The face of the picture tube now appears specked with minute points of light that endure for a fleeting instant and then disappear. This sparkling causes the observer evaluation to flatten out above 4.5 mc. It occurs only when the total noise energy is contained in the area above 4.5 mc. This effect appears related to the random oc-

Table IV. Computation of Weighting from Bandpass Data of Fig. 2

(1) Mid-band Frequency, Mc	(2) Full Band-width, Mc	(3) Total Relative Power for Equal Interfering Effect, Db	(4) Corrected Frequency
0.325	0.050	0	0.325
0.505	0.165	1.4	0.505
0.750	0.264	2.9	0.737
1.12	0.388	4.5	1.097
1.71	0.600	8.0	1.644
2.55	0.889	13.3	2.457
3.80	1.29	19.9	3.585
6.15	2.14	23.4	6.15

Table V. Computation of Weighting from Low-Pass Data of Fig. 3

(a) Bandwidth, Mc	(b) Bandwidth Excess Over Adjacent Lower Band	(c) 10-Log Excess Bandwidth 0.040	(d) Relative Power per 10 Kc for Equal Interfering Effect, Db	(e) Weighted Power in Adjacent Lower Band	(f) Total Weighted Power for Given Interfering Effect	(g) Weighted Power in Bandwidth in Excess of Lower Adjacent Band	(h) Total Level of Bandwidth in Excess of Lower Adjacent Band, Db	(i) Weighting for Bandwidth in Excess of Lower Adjacent Band, Db	(j) Mid-Frequency of Band in Excess of Lower Adjacent Band, Mc
0.040			0		1.000		0	0	0.020
0.080	0.040	0	−2.55	0.525	1.000	0.475	− 3.2	− 0.65	0.060
0.133	0.053	+ 1.2	−3.77	0.756	1.000	0.244	− 6.1	− 3.53	0.106
0.211	0.078	+ 2.9	−5.27	0.708	1.000	0.292	− 5.3	− 2.93	0.172
0.320	0.111	+ 4.4	−5.72	0.902	1.000	0.098	−10.1	− 8.78	0.266
0.585	0.265	+ 8.2	−6.74	0.791	1.000	0.209	− 6.8	− 8.26	0.452
0.890	0.305	+ 8.8	−7.11	0.918	1.000	0.082	−10.9	−12.59	0.737
1.360	0.470	+10.7	−7.53	0.908	1.000	0.092	−10.4	−13.57	1.125
2.300	0.940	+13.7	−7.56	0.993	1.000	0.007	−21.5	−27.64	1.830
3.075									
4.73									
7.22									

currence of peak potentials in the random noise source and to the effect of these potentials on the kinescope. The output of the noise generator was examined periodically with an oscilloscope for evidence of clipping. These measurements indicated the output of the noise generator to have a peak factor of 14 db (peak noise volts to rms noise volts) and showed no evidence of clipping.

Overloading of the amplifiers external to the viewing monitor, Fig. 2, tended to change the appearance of the sparkling. The minute bright areas became elongated and more nearly resembled short dashes of light. This change in character made it possible to differentiate between the true sparkle effect and overloading of the amplifiers. Three different *1816 P4* kinescopes showed the sparkle effect. This suggests that the effect is characteristic of the kinescope.

The interpretation of this result on design standards is of importance. If a situation arises in which the expected noise is all in the region above 4 mc, the upper weighting curve should be employed, since the noise cannot be permitted to break through the threshold of the sparkle effect. If the expected noise is spread out over the low as well as the high frequencies, it would be permissible to use the lower curve, since the contribution of the high frequencies (when they are below the threshold of the sparkle effect) is negligibly low and is probably most accurately assessed by application of this weighting. As a practical matter, on general types of noises, it would probably not matter what curve is used, since the sparkle effect begins at levels over 25 db higher than the threshold of interference in the low-frequency regions.

Two-Band Noise Tests

Each 2-band noise was made up of a low-pass band (cutoff low on the frequency scale) and a high-frequency bandpass band. The levels of each were adjusted to be equally interfering. The observations were made in the same manner employed for the 1-band tests described earlier and shown schematically in Fig. 2. Ten observers were employed. Table III gives the results of these tests and indicates the weighted power in the 2-band noises were from 1 to 1.4 db smaller than the weighted power in the judgment criterion. These results substantiate the general results presented here as regards the magnitudes of the deviations expected. More specifically, the added bands fall in the category of narrow bands and thus should be underweighted by the compromise weighting, as shown in Figs. 6 and 7, and as is evident in the results.

Conclusion

The results of the experiments described have indicated that random television interference can be measured with a power meter having frequency weighting, with sufficient accuracy to be of value in the design and maintenance of television transmission circuits.

The relationship between the quantity measured and the subjective effect is not precisely unique but the order of magnitude of the errors is about the same for the weighted power meter as were the rms deviations of the observer data. Useful design and maintenance standards can be set up in terms of meter readings at given level points on transmission circuits.

Appendix. Weighting from Bandpass and Low-Pass Data

Bandpass Data

The derivation of a frequency weighting from the bandpass data shown in Fig. 3 is given in Table IV. In the table, column 1 shows the mid-frequencies of the respective bands, and column 2 the full bandwidths. Column 3 gives the total power in each band in db relative to the power in the first band when each noise was adjusted to have the reference interfering effect. In column 4 corrected frequency figures are given. These figures are very nearly the same as the figures in column 1 (mid-band frequencies) but slightly different because, if an appreciable slope in the weighting occurs over one of the given bands of noise, it is apparent that the center-of-interference of that band is not the arithmetic center of the band. Since it was desired to plot the weightings at the center-of-interference of each band, a small frequency correction was calculated whenever the db weighting across the band changed by 2 db or more. The figures in columns 3 and 4, when plotted, give the desired shape of the weighting. This is the shape given in the top curve of Fig. 6. As indicated in the text, the position of the curve on the db scale was determined by adjustment, to produce calculated magnitudes of noise from the bandpass data that were equal to the magnitudes of the low-pass noises calculated from the bottom curve.

Low-Pass Data

The calculation of weighting from the low-pass data yields results which are considerably more inaccurate than those obtained from the bandpass data, especially in the frequency regions above 500 kc. This is because all the judgments from which weightings may be calculated were made in the presence of low-frequency noise which has been shown to be more interfering, per cycle, than high-frequency noise. Hence, small differences in the opinions of observers in the levels producing equal interfering effects result in very large differences in the estimated weightings for the frequency regions above the very lowest employed.

A very effective method of weighting derivation is the cut-and-try method. This, however, usually involves several unsuccessful attempts and is correspondingly tedious because of the computations involved. Hence, a method of obtaining an initial curve from the data was worked out and a sample is carried out in Table V for the data given in Fig. 4.

The principle involved in making the calculation was that each noise should be required, through the application of a calculated weighting, to produce a weighted power sum of unity, since each was judged to be equally interfering. A weighting of 0 db was assumed to apply all across the lowest band (40 kc wide). In the next wider band a 0-db weighting was assumed to hold over that portion common to the lowest band, and a lesser weighting calculated for the remainder of the bandwidth such that the total weighted sum of the power for the noise came out to be unity also.

For the third band the portion common to the second band was effectively treated with the two weightings previously derived and the remainder, in excess over the width of the second noise, was operated upon by a third weighting calculated to make the weighted sum of the powers unity. This method was carried on for bands up to 2.3 mc wide. For wider bands the calculation indicated that practically no contribution to interference was being made by the high frequencies and hence no evidence could be adduced as to where the weightings should be.

In Table V, column a gives the bandwidths of the low-pass noises employed. Column b gives the excess bandwidth, for any particular low-pass band, over that of the adjacent lower band. Column c gives the frequency factor in db, of the excess bandwidths given in column b. Column d gives the relative powers per 10 kc in db for the bands when they were adjusted in level to produce equal interfering effects. These data are obtained from Fig. 4, making the db figures relative to that for the narrowest band (40 kc). Column e gives the weighted power in that portion of any band common to the next lower adjacent band. These are power units and are obtained by reducing unity by the difference between the db figure in column d and the corresponding figure in the row immediately above. For example, the figure 0.525 in column e corresponds to unity reduced by 2.55 db. Column f gives the power figure of unity to which the calculation in each row is directed. Column g gives the difference between the figures in columns e and f. Column h is the db figure corresponding to the power figures in column g. Column i gives the desired weighting. This figure is obtained by combining the figures in columns c, d and f according to the equation $i = h - (d + c)$.

Column j gives the frequencies corresponding to the weightings. These frequencies are approximately the mid-frequencies of the portion of a given low-pass band not common to the next adjacent lower band.

The smoothed curve drawn through these points, averaged with a similarly derived smooth curve for the data shown in Fig. 5 gives the shape of the bottom curve shown in Fig. 6.

References

1. DATA ON RANDOM NOISE REQUIREMENTS FOR THEATRE TELEVISION, **Pierre Mertz.** *Journal*, Society of Motion Picture and Television Engineers, New York, N. Y., vol. 57, Aug. 1951, pp. 89–107.

2. IMAGE GRADATION GRAININESS AND SHARPNESS IN TELEVISION AND MOTION PICTURE SYSTEMS, PART 2, **Otto H. Schade.** *Journal*, Society of Motion Picture and Television Engineers, New York, N. Y. vol. 58, March 1952, pp. 181–222.

No Discussion

Fast Response with Magnetic Amplifiers

D. G. SCORGIE
MEMBER AIEE

FOR MANY years the use of magnetic amplifiers was confined to those applications where extreme reliability was required, but where speed of response was not particularly important. Under the pressure of greater reliability and weight-saving requirements in higher power installations, important advances have been made in magnetic amplifier circuitry and components to achieve higher gain-to-time-constant ratios. It was not until the work of R. A. Ramey,[1] however, that attention was focussed on the fact that magnetic amplifiers could be made with extremely fast response (i.e., one-half cycle of the power frequency) more or less independently of the power gain.

Self-saturating magnetic amplifiers, in which positive feedback is achieved by passing the rectified load current through the output windings, are known to give high gain-to-time-constant ratios. Recent work[2,3] has shown that, with suitable combinations of external positive and negative feedback, it is possible to achieve extremely high input impedances when using the self-saturating magnetic amplifier. W. A. Geyger has shown how to increase the speed of response considerably by delaying the negative feedback through the use of linear low-pass filters.[2]

With a view toward achieving extremely fast response without sacrificing power gain, the present paper will explore the use of positive and negative feedback in self-saturating circuits. Half-cycle response is attained without recourse to rectifiers in the control circuit, thus in many cases removing a lower level limitation upon the signal voltage. The half-cycle response amplifiers to which we refer reach 100 per cent of their final steady output one-half cycle after a transient signal voltage is applied.

Analysis of an Idealized Self-Saturating Amplifier

The conventional bridge-connected self-saturating magnetic amplifier is shown in Fig. 1. The symbol designating the wire-wound cores shows a rectangle, signifying the idealization of a perfectly square hysteresis loop core material. For simplicity we call the element a remanator because it is characterized by 100-per-cent remanent flux. The amplifier shown in Fig. 1 has been discussed at length in the literature.[4,5] Seemingly, it would be needless repetition to review its operation again. However, certain of the results achievable depend upon the adoption of a new point of view. Ramey[1] attributed his results to "a recognition of the fact that the magnetic amplifier is a voltage-sensitive device and not, as generally believed, a current-sensitive device." Taken from context, the statement has very little consequence, for surely Ohm's law holds, and the control current and voltage are linearly related as determined by the control circuit resistance. However, when we perform the thought experiment of setting the control circuit resistance R and voltages e_c and e_b, Fig. 1, equal to zero, the control circuit current i_c assumes various values about which Ohm's law can provide us with no information. Therefore, this thought experiment is an appropriate beginning from which to exhibit the dependence of the control current upon the magnetic amplifier operation, rather than vice versa.

With the terminals of the control circuit windings short-circuited and zero-winding resistances, it follows that the winding voltages must be instantaneously equal and opposite. On the half-cycle shown in Fig. 1, core I is the gate, and it will absorb a certain voltage-time integral from the supply prior to saturation.

Paper 53-336, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Pacific General Meeting, Vancouver, B. C., Canada, September 1–4, 1953. Manuscript submitted June 4, 1953; made available for printing July 15, 1953.

D. G. SCORGIE is with the Naval Research Laboratory, Washington, D. C.

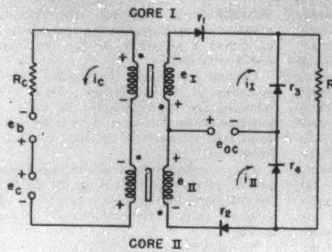


Fig. 1. Conventional bridge-connected self-saturating magnetic amplifier

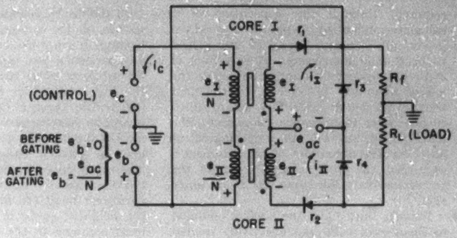


Fig. 2. Amplifier with full response in one-half cycle of power supply

After the saturation of core I, the supply voltage, e_{ac} is delivered to the load. Because of the short-circuited control windings, the voltage e_{II} on core II is of necessity equal to e_I, resulting in a reduction of the core-II flux from its fully saturated condition. The same operation is repeated on subsequent half-cycles, with the same flux reduction being transmitted to the resetting core. Since the initial flux reduction and supply voltage completely determine the output voltage, a steady-state average output current must flow. For a fixed supply and load, the value of this average output current depends only upon the history of the cores. Specifically, it depends upon the initial flux reduction of core I on the first half-cycle and the assumption that core II was initially saturated.

No mention of control currents was necessary above in showing how the various outputs arose. The voltage supply e_{ac} was the only source of power. e_{ac} and the physical circuit, including the history of the cores, will completely determine the dependent control winding current. By definition, the control current i_c flows through the signal element. In other circuitry i_c is not necessarily the same as the control winding current, although of course these currents are identical in Fig. 1.

During the period of time prior to saturation of the gating core the instantaneous magnetizing current i_m of the resetting core must be such as to allow the transfer of the volt-second integral from one core to the other. In this interval the currents will be determined primarily by core materials and output rectifier leakage. After firing, or gating, however, if we assume core materials with a square hysteresis loop, the control current would drop to zero unless back leakage through rectifiers r_1 and r_2 were large (the saturation current I_s or more). In general I_s is defined as that magnetizing current which just saturates the core. For perfectly square loop material, this is the half width of the hysteresis loop and is equal to the instantaneous magnetizing current i_m. Using square-loop cores with the control windings short circuited and no independent control over the control current, the average control current, i_c is determined by the percentage of the cycle during which saturation current I_s is allowed to flow. That is

$$i_c = I_s \frac{\theta_f}{\pi} \tag{1}$$

where θ_f is the angle at which gating occurs. The average output current I_L and firing angle θ_f are related cosinusoidally

$$I_L = \frac{\hat{E}_{ac}}{\pi R_L}(1+\cos\theta_f) \tag{2}$$

so that

$$I_L = \frac{\hat{E}_{ac}}{\pi R_L}\left(1+\cos\pi\frac{i_c}{I_s}\right) \tag{3}$$

If, through a separate bias winding, a percentage, $1-K$, of the saturation current is passed prior to gating, then the right-hand side of equation 1 is multiplied by K and equation 2 becomes

$$I_L = \frac{\hat{E}_{ac}}{\pi R_L}\left[1+\cos\pi\left(\frac{i_c/K}{I_s}\right)\right] \tag{4}$$

The control current necessary to achieve the full range of output values is therefore made, by adding bias, as nearly zero as we wish.

Passing from one steady state to another was seen to be impossible with the control windings short circuited. However, if suddenly e_c of Fig. 1 is applied, the resetting core will no longer change level of induction at as rapid a rate as the gating core, although with no bias, the control current remains unchanged at I_s prior to gating. After gating the control current reverses to $(-I_s)$ and the direction of change of flux level in the resetting core reverses again toward saturation. Over a period of several cycles, the change of output will be proportional to the integral of control voltage (integrated over the several cycles). The integration process would continue to saturation output unless the control voltage were removed, leaving matters in some steady state described by equation 3. This idealized arrangement is an integrator and is useless as an amplifier. Yet we see that due to the integration action very little power was necessary to change from one steady state to another. Practical considerations enter the actual physical circuit to govern the extent to which the integration process is allowed to proceed, hence making the circuit controllable with a steadily applied input.

But first let us continue a bit further with the idealized circuit with perfect rectifiers and perfectly rectangular hysteresis loop core material. This time include resistor R_c in the control circuit along with voltages e_b and e_c. We saw the circuit before as an integrator of the control voltage. Now we modify matters by saying that any change of the circuit output will be the integral of whatever voltage it sees looking back from the control winding terminals, thus including the i_cR_c voltage drop and e_b in addition to e_c. A steady state is therefore defined when

$$\int e_c dt + \int i_c R_c dt - \int e_b dt = 0 \tag{5}$$

integrated over each half-cycle of supply frequency. That is, of course, self-evident in the equivalent statement that a steady state exists when

$$\int N_I \frac{d\phi_I}{dt} dt + \int N_{II} \frac{d\phi_{II}}{dt} dt = 0 \tag{6}$$

integrated over each cycle. (ϕ_I and ϕ_{II} are the fluxes linking the control windings of cores I and II respectively.)

The statement of control, equation 5, was set up in this fashion to emphasize two modes of control which are significantly different.

First is the usual case where e_b and R_c are constant. When a change of control voltage e_c is introduced after equation 5 has been first satisfied, a steady state is impossible until the average I_c has changed enough to balance equation 5 once again. The natural inference has been drawn that the current I_c is the controlling variable.

Curiously enough, the idealized circuit of which we are speaking is still not controllable with a steadily applied control voltage. With cores of rectangular hysteresis loop of half width I_s, $e_b=I_sR_c$ and $e_c=0$, equation 5 balances when gating never occurs and the output current is zero. If e_c is applied, equation 5 can balance only after gating. Prior to gating $i_c=I_s>\left(\frac{e_b-e_c}{R_c}\right)$ so that equation 5 cannot balance until full output current is reached. The transient will be similar to an exponential rise, since the time interval prior to gating gets shorter, contributing less to the increase of output each half-cycle.

Any form of negative feedback, which either increases e_b or decreases i_c with output current, will stop the output from running to saturation. If, as in practical core materials, the sides of the hysteresis loop are not perfectly vertical, but have some finite positive slope, then the average value of the magnetizing current I_m of the resetting core decreases as gating comes earlier in each cycle. Therefore, it constitutes the additional negative feedback term in equation 5 necessary for control.

If, secondly, however, we allow R_c to be negligible and instead vary e_b following a change of control voltage, equation 5 can be easily balanced. In this mode of operation, when predicting the output current, what happens to the control current is a matter of complete indifference. Positive and negative current feedback do not affect the control characteristics. Voltage feedback, as manifest in e_b, completely determines control.

Current feedback and bias, taken together, are useful in reducing the control current, hence the control power, to a minimum throughout the control range. Notice particularly, however, that negative voltage feedback also plays a role in minimizing the a-c components of control currents. Following gating in the amplifier of Fig. 1 (with R_c small) the control current is no longer forced to its upper limit I_s, imposed by the control winding of the resetting core. If e_b, the negative voltage feedback, is introduced after gating, and is instantaneously greater than e_c, it will force I_s to continue to flow, still limited by the resetting core. With perfectly square loop cores, therefore, the control current will differ from the constant I_s only at the instant when the supply voltage passes through zero. It is easy to reduce this form of current to nearly zero with d-c bias, since it is relatively free of harmonic components.

Transient Response

The idealization of components in the amplifier as discussed in the foregoing gives an unrealistic picture of performance when applied to practical circuits. We were led, for instance, to the conclusion that the amplifier with square loop core material, operating in its normal mode, had virtually infinite power gain. Recitation of the departures from idealized components and the study of their effects are extremely important but will more properly be the subject of another paper. The main theme of this paper is the speed of response and improvement of power gain.

It has been argued in the foregoing that a change from one steady-state average output voltage to another is proportional to the left-hand side of equation 5, integrated over the intervening time interval. If we assume the output voltage to be that impressed across the whole output circuit resistance, this constant of proportionality is simply the turns ratio N between the output and control winding on each core. The output is cyclically pulsing. If a change of control voltage e_c' is impressed at $t=0$, then during the ensuing transient, changes of the amplifier variables (expressed as primed quantities) are related as shown by equation 7

$$e_o'(t_n)=N\left[\int_0^{t(n-1)}e_c'dt+\int_0^{t(n-1)}i_c'R_cdt-\int_0^{t(n-1)}e_b'dt\right] \quad (7)$$

$e_o'(t_n)$ is the net difference between the volt-second area of the output voltage during the nth half cycle, which ends at the instant t_n, and that which ended at $t=0$.

Dornhoefer[1] states with reference to the gating core that "an assumption that the voltage-time integral absorbed by the reactor is given by the integral of the supply voltage from zero (i.e., beginning of nth gating half-cycle) to the firing instant is a relatively good one." The voltage-time integral that the reactor was capable of absorbing was, of course, determined during the previous $(n-1)$ half-cycle during which the same core was reset. The change of this latter quantity during a transient is described by the three terms on the right-hand side of equation 7. If the first term is positive, the others are normally negative quantities which gradually increase in magnitude as some function of the output voltage pulses. As Geyger points out, the response is faster if the buildup of the last two terms of equation 7 is delayed.[2] One might go further and suggest that it is possible to speed matters further by briefly introducing the last term of equation 7 as a positive quantity.

We will inquire here into the specific case where a step function of e_c' is applied at $t=0$, and the second and each succeeding half-cycle voltage $\int e_b'dt=\int e_c'dt$ is fed back from the output circuit without changing i_cR_c. This is a sufficient condition to assure an amplifier with full response in one-half cycle of the supply frequency. Direct application of equation 7 shows that for the first half-cycle $e_o'(t_1)=0$. On the second and every succeeding half-cycle $e_o'(t_n)=N\int_0^{t_1}e_c'dt$, so that an amplifier with half-cycle

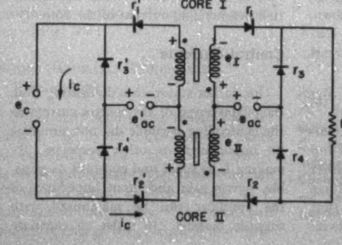
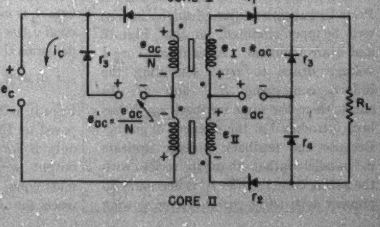


Fig. 3 (left). Ramey double-bridge self-saturating magnetic amplifier

Fig. 4 (right). Equivalent of Fig. 3 during resetting of core II



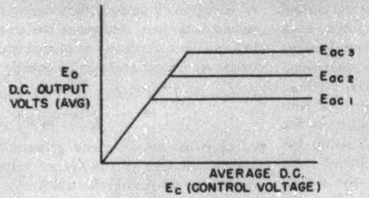


Fig. 5. Voltage transfer characteristics, circuits of Figs. 2 and 3

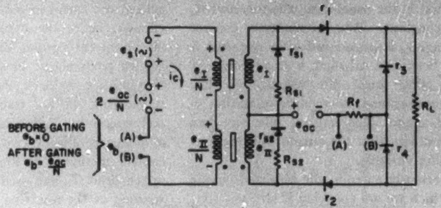


Fig. 6. A-c-controlled amplifier with one-half cycle response and low quiescent output

response is obtained. The power gain is determined only by how effectively control current can be reduced to a minimum by bias and current feedback.

D-C-Controlled Amplifiers with Half-Cycle Response

Fig. 2 shows the simplest embodiment of an amplifier which responds in the manner described in the previous paragraph. The quantity e_b is simply derived from the load by feeding back approximately e_{ac}/N to the control circuit instantly upon saturation of the gating core. The simplest method is the direct coupling through ground and terminal, A of Fig. 2. The circuit diagram shows the use of remanators, or square loop core materials. Theoretically the response time and voltage transfer characteristics are equally satisfactory with any saturable magnetic material, because there is nothing in the control circuit to upset the required voltage relationships when the magnitude or direction of the control current changes. The substitution of various core materials will alter only the control current-output voltage relationship, as discussed later.

Fig. 3 shows the double-bridge self-saturating circuit of Ramey,[1] which also responds fully in one half cycle of the supply frequency. Using remanators for cores I and II, the feedback action is identical to that of the circuit of Fig. 2, although not readily apparent. While core I is on its gating half-cycle and core II is resetting, the effective portion of the double-bridge circuit is shown in Fig. 4. Prior to gating rectifier r_2' might just as well be open-circuited, so that Figs. 2 and 4 are identical. After firing, core I voltage drops to zero, activating the branch comprising E_{ac}' and r_2'. In effect, then, the same feedback e_{ac}/N has been introduced at the moment of gating. Because the feedback voltage appears in parallel rather than in series with the gating core, rectifier r_1' is necessary to prevent destructive currents from flowing in the control winding after gating. The double bridge, Fig. 4, requires remanators for satisfactory operation.[6] However, with additional circuitry its use has been shown to extend to other core materials.

When properly designed, the voltage transfer characteristics of the circuits of Figs. 2 and 3 have the identical form for pulsating d-c (rectified sinusoid) control, as shown in Fig. 5. For pure d-c control, the transfer characteristics of the circuit of Fig. 3 are altered somewhat but usually not to an important extent. The linear portion in any case is practically independent of supply voltage, but in the saturation (horizontal) portion, the output voltage is of course directly proportional to the supply voltage.

A-C Controlled High-Speed Amplifier

The circuit of Fig. 6 without the shunting elements R_{s1}, R_{s2}, r_{s1}, and r_{s2} is an a-c controlled self-saturating magnetic amplifier with an alternating voltage instantaneously equal to e_{ac}/N fed back to the terminals (A and B) in the control circuit at the instant load current flows. $2e_{ac}'$ is a fixed supply voltage equal to $2e_{ac}/N$. We find that especially where core materials with high flux remanence are used, there is a certain fairly large minimum output current for either polarity or any magnitude of the alternating control signal e_s. The reason for this high quiescent current can be seen by examining the voltage distributions in the circuit of Fig. 6 without shunts R_{s1}, r_{s1}, R_{s2}, r_{s2}. With the polarities as shown, core I is on its gating half-cycle. Kirchoff's law applied to the control circuit demands that

$$e_I+e_{II}=(2e_{ac}'-e_s)N=2e_{ac}-Ne_s \qquad (8)$$

Prior to saturation, the only limitation imposed by the output circuit is that if and only if forward current flows in either output winding, that winding voltage must equal e_{ac}. From this it follows that when, for any reason, forward currents do not flow in the output windings, there are many possible voltage values which satisfy equation 8. The only limits are

$$e_I \geqq e_{ac} \text{ and } e_{II} \leqq e_{ac} \qquad (9)$$

In practice, there are cumulative small effects such as unbalance of magnetizing currents, forward voltage drops in the output rectifiers r_3 and r_4, etc., which produce the inequality relationships of equation 9 at least a portion of the time when $e_s=0$. Here we see that if the elements R_{s1} (assumed that $R_{s1}>>R_L$) and r_{s1} are inserted when core I is on its forward half-cycle, the source e_{ac} delivers a small amount of power to R_{s1} and unequivocably fixes the voltage e_1 at e_{ac}. Therefore, under all conditions the resetting core voltage e_r must be unequivocably fixed at

$$(e_r)=(e_{ac})-(Ne_s) \qquad (10)$$

in the $(n-1)$ half-cycle.

In the nth half-cycle the area of the output voltage pulse $\int_n e_o dt$ is given by

$$\int_{t(n-1)}^{t_n} e_o dt=\int_{t(n-1)}^{t_n} e_{ac} dt-\int_{t(n-2)}^{t(n-1)} e_r dt \qquad (11)$$

From equations 10 and 11, and the assumption that e_{ac} is constant

$$\int_{t(n-1)}^{t_n} e_o dt=N\int_{t(n-2)}^{t(n-1)} e_s dt \qquad (12)$$

Equation 12 says not only that we have a linear magnetic amplifier which responds fully in one half-cycle of the supply frequency, but that the output (quiescent) is zero when e_s is zero. Experimentally, it is found that the quiescent output current is on the order of twice the saturation magnetizing current of the core.

Control Currents

Our earlier statements derogating from the effectiveness of the control current as the controlling variable did not minimize its importance in other respects. By making i_c as small and constant as possible, power gain and input impedance of the amplifier can be maximized with suitable biasing. In order to calculate

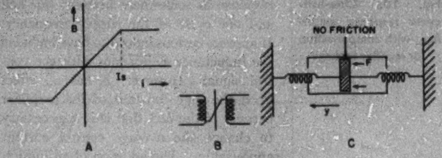


Fig. 7. A. Idealized saturable inductance characteristic. B. Symbol for saturable inductor. C. Mechanical analogy for saturable inductor

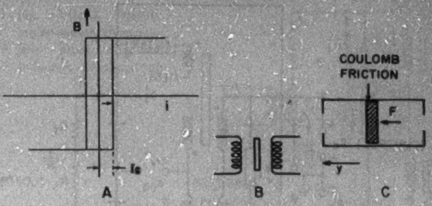


Fig. 8. A. Idealized remanator characteristics. B. Symbol for remanator. C. Mechanical analogy for remanator

transient response of the amplifier by equation 7 i_c must be known, unless $R_c=0$. The degree to which we can assume R_c negligible in the amplifiers with half-cycle response, is also dependent upon how successfully i_c is held constant. Of course R_c is always finite, so that the latter point is important.

The component which dominates the determination of control current is the saturable core and its magnetic material. To determine the current at a given instant one must know the limiting impedances and also be able to recognize the independent sources of energy. We know, of course, that e_{ac} and e_c represent sources. Fig. 7 shows idealized characteristics of saturable inductors, in which energy can be stored, and Fig. 8 shows those of remanators which are purely dissipative magnetic elements. The remanator is characterized by the high value of remanent induction which gives it a square hysteresis loop. We can get a mechanical analogue for both a remanator and saturable inductance with a weightless piston sliding in a cylinder with mechanical stops at both ends. Plot the linear displacement from the center as the vertical or y axis and the force F applied to the piston on the horizontal axis. With no friction and a spring attached to the piston, the saturation characteristic is identical to Fig. 7(A). When the spring is stretched energy $\int \mathbf{F} \cdot d\mathbf{y}$ is stored (as is $\int \mathbf{H} \cdot d\mathbf{B}$ for the magnetic case) and can be recovered by releasing the piston. Detach the spring and introduce coulomb friction, and saturation characteristic will change approximately to that of dissipative remanator of Fig. 8.

WHEN SATURABLE CORES ARE REMANATORS

If perfectly square loop remanators are used in Fig. 2, the control current takes an extremely simple form. When core I is on its gating half-cycle, as shown, core II is completely passive. That is, it can deliver no energy. The dominant source of energy is e_{ac}; core II and rectifier r_2 are the principal elements limiting control current. Prior to firing, e_{ac} is impressed across the output windings of core I, and e_{ac}/N appears across its control windings. Assuming $(e_{ac}/N)>e_c$ the control current i_c rises until limited by the resetting core II. The limitation imposed by core II is that the sum i_c+Ni_{r2}, where i_{r2} is the leakage current of rectifier r_2, be equal to I_s (the half width of the hysteresis loop). After gating e_b becomes the primary driving source and core II, as its resets, continues to limit the control current. Fig. 9(B) shows the control current for all outputs in the case where rectifiers r_1 and r_2 have infinite back resistance. The control current will be reduced by rectifier leakage. The voltage Ne_c is impressed instantaneously in the back direction across r_1 or r_2, whichever core is resetting. Therefore as E_c (hence output current) gets larger, the average control current decreases somewhat from its initial value of the magnetizing current I_s.

The simplicity of the form of the control current shown in Fig. 9(B) suggests that its reduction is not difficult. The major portion of the control winding current can be prevented from passing through the control source simply by shunting the control by a constant current source, I, which delivers I_s constantly. Fig. 10 is the circuit of Fig. 2 with this constant current source and an auxiliary winding added. The auxiliary winding is to compensate for the decrease in control winding current as output rectifier leakage increases. The auxiliary winding can also be shown as a few turns in series with the d-c load, in which case, however, smoothing with a filter is more difficult. When low quiescent current is important, it is often necessary either to insert an additional d-c bias voltage in series with the control windings to compensate for I_s flowing through the control winding resistance,

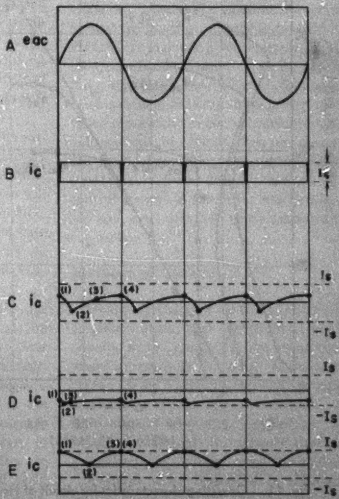


Fig. 9. Variables in circuit of Fig. 2. A. Supply voltage. B. Control current, all outputs, using remanators. C. Control current using saturable inductors, intermediate output. D. Control current using saturable inductors, nearly full output. E. Control current using saturable inductors, zero output

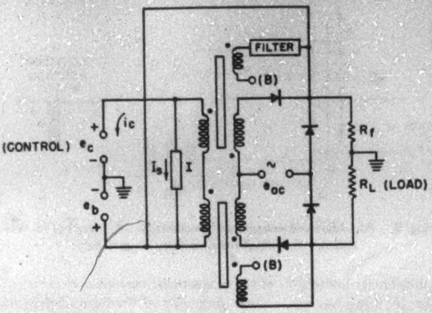


Fig. 10. One-half-cycle response amplifier, with compensation for rectifier leakage

or to allow I_s to flow through separate windings.

Fig. 11 shows a typical set of curves taken on an amplifier (set for one half cycle response) which has a turns ratio, (output winding)/(control winding), of 5,700/1,540 or nearly 4. Deltamax 1-mil toroidal cores were used. Where thicker magnetic tape is used the bowing of the control current versus output current curve is more pronounced because of eddy currents in the core. It does not necessarily follow that improved over-all performance can be obtained by using still thinner grain-oriented nickel-irons, for a deterioration of the high remanance characteristics accompanies further rolling. However, it could be that other materials such as the newly developed Alfenol with its high electrical resistance and rectangular hysteresis loop, or 4-79 molybdenum Permalloy, for which the B_r/B_s ratio increases in the thinner tapes, will provide more satisfactory results.

In supermalloy the B_r/B_s ratio is on the order of 0.7. However, satisfactory results can still be obtained with its use because of the extremely small coercive force. Fig. 12 shows the characteristics of a low-level amplifier adjusted for full response in somewhere between one-half and one cycle of the supply frequency. It appears curious indeed to have included the inductance L in series with the control signal. However, L has no effect upon the time constant of the amplifier because of the fact that it is unnecessary to change the average control current appreciably. The addition of an inductance has the effect of reducing the harmonic currents which must flow through the control. It is particularly effective in reducing the average control current in the region of low output currents.

The control characteristics of Fig. 12 illustrate particularly well the fact that, at least in some instances, a magnetic amplifier must be considered as a voltage-controlled device. The output current is seen to be a many-valued function of the control current, but is stable and uniquely determined by average control voltage. The control current transfer curve has a characteristic S-shape when positive feedback is used. The central portion can be made nearly vertical by decreasing the positive feedback, but the curve at the bottom of the S remains unchanged. More accurate compensation can be effected by introducing suitable nonlinearities in the positive feedback loop. However, it is more satis-

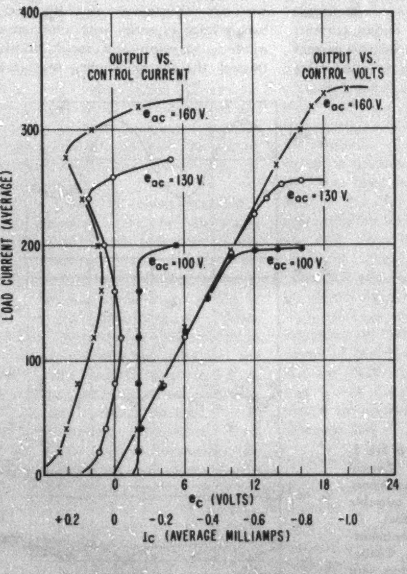


Fig. 11. Operating characteristics for circuit of Fig. 10

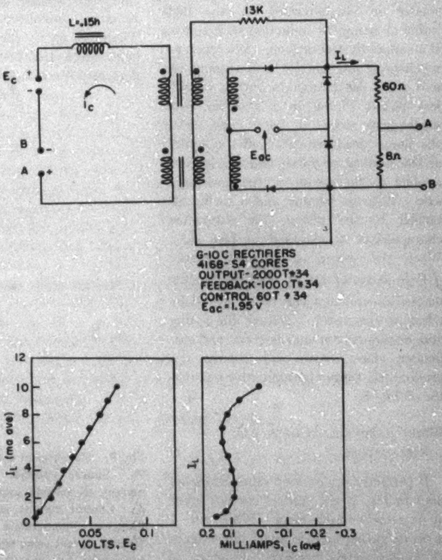


Fig. 12. Operating characteristics for low-level high-speed magnetic amplifier

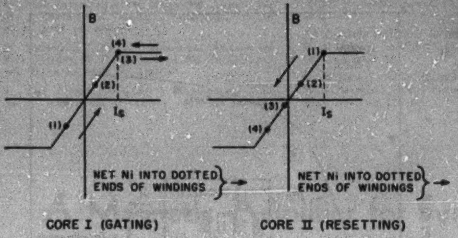

B
(4)
(3)
(2)
(1)
I_S
NET NI INTO DOTTED ENDS OF WINDINGS
CORE I (GATING)
B
(1)
(2)
(3)
(4)
I_S
NET NI INTO DOTTED ENDS OF WINDINGS
CORE II (RESETTING)

Fig. 13 (left). Traversing saturation characteristic using saturable inductors in circuit of Fig. 2

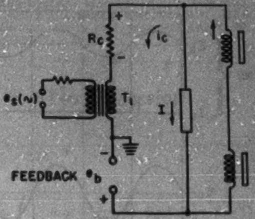


Fig. 14. Control circuit of Fig. 10, modified for driving a 2-phase a-c motor

factory to introduce means of reducing the average control current in the region of low outputs as, for instance, with the series inductance L mentioned in the foregoing.

The reason for the effect of L upon the average control current is that the a-c components of control current must pass through terminals A and B, which are conductively coupled to the output rectifier bridge. The slight rectifying action, which is greatest for low output currents, gives rise to a d-c component of voltage at terminals A and B. Thus a decrease in a-c component control current gives rise to a decrease in the d-c component. Three additional factors contribute to the curvature of the control current characteristic in the region of low outputs.

Unbalance Between Cores

The output current is reduced by increasing the voltage impressed upon the resetting core. Ideally, the resetting core voltage should be instantaneously equal to e_{ac} for zero output. In practice, however, a slight difference between the turns ratios of the two cores brings about a condition where the resetting core voltage tries to rise above e_{ac}. As a consequence, a small current flows in the forward direction through the output rectifier r_1 or r_3 in series with the resetting core. This occurs at a time when the resetting core is limiting the control winding current, so that i_c rises by the same amount multiplied by the turns ratio.

Excess of Negative Feedback Voltage

At the instant of gating, the voltage e_{ac}/N is fed back into the control circuit. If too great a feedback pulse is introduced, particularly when the control voltage is small, the resetting core voltage rises above e_{ac}. The limitation on the control current imposed by the resetting core is then removed just as when the cores were unbalanced.

At the full output end of the characteristic, Fig. 12, the control current drop sharply. This can be improved by increasing and linearizing the back resistance of the output rectifiers by derating. However, the basic difficulty remains that the core materials are never possessed of perfect remanence. The pulse of energy delivered by the core as it begins to reset decreases the control current. The effect becomes more important percentagewise when we wish to prevent reset altogether and produce full output.

Excess Supply Voltage

Where the supply voltage is sufficiently high to swing the flux level of the cores from one knee to the others in a single half-cycle, there is the danger of saturating the resetting core. This removes the current limitation imposed upon the control current by the resetting core. The situation is corrected by decreasing the supply voltage.

WHEN CORES ARE SATURABLE INDUCTORS

Once again, we will determine the control currents in the circuit of Fig. 2. This time replace cores I and II by the idealized saturable inductances of Fig. 7, wherein energy can be stored. We can no longer assume that the supplies for e_c and e_{ac} are the only sources of energy; instead one or the other or both of the cores will in general also be delivering energy.

Fig. 13 shows the traversal of the saturation curve of the saturable inductors during a single half-cycle of operation. Assuming that the magnetic induction will swing from knee to knee with the application of the full supply voltage, the case shown would represent the condition of about one-third maximum output current.

The half-cycle begins at point 1, Fig. 13. At this point the resetting core has its maximum stored energy. In an effort to dissipate this energy, it raises the voltage e_{II} to the value of e_{ac} and causes a small current i_{II} (Fig. 2) to flow in the output circuit. The control voltage is added to the resetting core voltage, so that the gating core voltage rises to $e_I = e_{ac} + Ne_c$. Thus rectifier r_1 will block and no current flows in the output winding of core I. In direct contrast to the case where remanators were used, we now see that core II is the primary energy source, and core I the limiter of control current. Between points 1 and 2 of Fig. 13, the control current is determined by core I, and it changes at a rate proportional to $(e_{ac} + Ne_c)$. Point 2 of Fig. 13 is reached at the instant the total magnetizing current requirements of the two cores become the same. The current i_{II}, which has been diminishing, disappears and i_I begins to flow in the output winding of the gating core. At point 2 it can be seen that i_c has reversed polarity Fig. 9(C). Beyond point 2, the situation reverts to that which was normal with remanators, namely, the resetting core limits the control current, $e_I = e_{ac}$, and $e_{II} = e_{ac} - Ne_c$. The rate of change of i_c is proportional to e_{ac}. At point 3 gating occurs, but i_c is unaffected and continues at the same rate until the end of the half-cycle, point 4.

Fig. 9(C) showed an example of control current with some intermediate output. Figs. 9(D) and 9(E), respectively, show the control current wave shape for nearly full output and zero output. The contrast is evident between this and Fig. 9(B) which represents the control current for all values of output current when remanators are used. Preventing the control winding currents from flowing through the control sources is obviously less formidable and more effective when remanators are used.

Methods of Control

Unique methods of control of magnetic amplifier have been proposed.[7] Their novelty rested chiefly on the point that where remanators are used, the control current flows backward through the control voltage source, Figs. 2 and 9(C). A fixed amount of power, supplied by the voltage source e_{ac}, is transformed through the gating core and delivered to the control circuit, where it must all be absorbed by the control element and the resetting

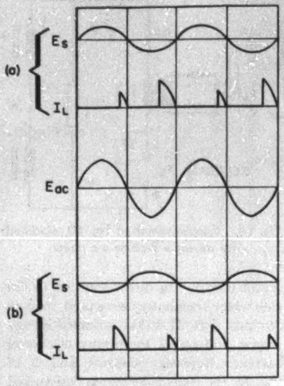


Fig. 15 (left). Output current I_L, when control voltage e_s, is introduced as shown in Fig. 14. A. Forward phase of e_s. B. Reverse phase of e_s

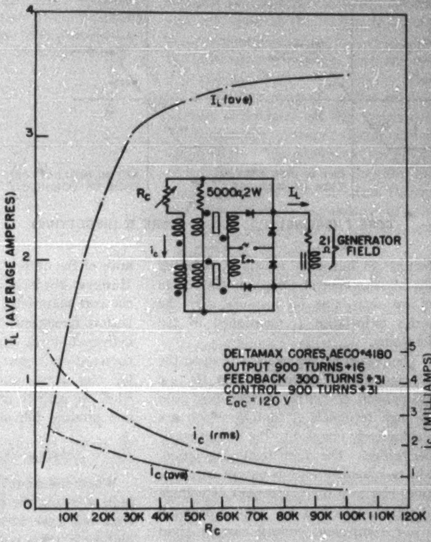


Fig. 16 (right). Operating characteristics of high-speed resistance-controlled amplifier with inductive load

core. Increased output is obtained by absorbing a greater percentage of this power in the control element. Therefore, any variable dissipative element can be used for control.

We shall discuss two examples, the first illustrating the utility of combinations of passive and active control elements.

Fixed Resistor Plus Variable Voltage Control

Resistance control has been discussed by Ramey.[7] Subsequently, Geyger showed the advantages of using superimposed a-c and d-c components in the output of a magnetic amplifier to drive a 2-phase motor with poor damping characteristics.[8] Lufcy, Schmid, and Barnhart disclosed circuitry in which a fixed resistor and variable control voltage were combined to control a high-speed magnetic servo amplifier.[9] The combination of the two central ideas[8,9] is achieved as shown in Fig. 14. There, only the control circuit of Fig. 10 is shown, wherein the control voltage e_c has been replaced by the combination of resistor R_c and a transformer T_1 through which an a-c signal of reversible phase is introduced.

Fig. 14 shows a constant current source I which, if included at all, must be set at some value less than the saturation magnetizing current I_s of the remanator control windings. A unidirectional current flows through R_c giving an effective d-c control voltage. The value of R_c is chosen so that this effective control voltage gives about one-half full output from the amplifier. When the alternating signal e_s is introduced, it alternately adds to and substracts from the voltage across R_c on successive half cycles. The value of i_c is unchanged except as e_s effects output rectifier leakage or eddy currents in the resetting core. Fig. 15 illustrates how a change in phase of e_s brings about a reversing phase fundamental a-c component in the amplifier output.

Resistance-Controlled Amplifier with Inductive Load

This example is included to illustrate several points which have not been specifically mentioned. Chief among these is a reaffirmation of the proposition that amplifier performance can be profitably analyzed in terms of voltages as the principal variables. However, in this instance the statement applies to the output circuit as well as to the control.

Fig. 16 shows the transfer characteristic of an amplifier which is controlled by a variable resistor R_c and has a highly inductive load. The load voltage responds fully, one half cycle after a step change of control resistance. However the load current rises exponentially with a time constant determined by its inductance and resistance. The negative voltage feedback pulses which have been used throughout the paper to achieve half-cycle response cannot be obtained in this example from a resistor in series with the load. It would be necessary to provide an instantaneous voltage divider across the load terminals for feedback if the output-to-control winding turns ratio N were greater than one. With resistance control the turns ratio can be one and the whole load voltage acts as the feedback voltage E_b. In practice the turns ratio N should be greater than one because of the internal resistance of the output windings.

Summary

The bridge-connected self-saturating magnetic amplifier can be made to respond fully to a signal in one half cycle of the supply frequency. A number of types of input have been considered: a-c, d-c, and combinations of the two for performing a variety of functions. Analysis of the circuits, assuming either current or voltage as the control variable, is useful, but examination of the control voltages is generally the more so. Remanators, in which no energy can be stored, lead to circuitry in which most of the control winding current can be prevented from passing through the control source.

References

1. On the Mechanics of Magnetic Amplifier Operation, R. A. Ramey. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1214–23.

2. Magnetic Amplifiers of the Self-Balancing Potentiometer Type, W. A. Geyger. *AIEE Transactions*, vol. 71, pt. I, 1952 (Jan. 1953 section), pp. 383–95.

3. Magnetic Amplifier, A. S. Fitzgerald. United States Patent No. *2,464,639*, 1949.

4. Self-Saturation in Magnetic Amplifiers, W. J. Dornhoefer. *AIEE Transactions*, vol. 68, pt. II, 1949, pp. 835–50.

5. Response Time of Magnetic Amplifiers, E. L. Harder, W. F. Horton. *AIEE Transactions*, vol. 69, pt. II, 1950, pp. 1130–41.

6. The Effect of Core Materials on Magnetic Amplifier Circuits, Leo J. Johnson. *AIEE Transactions*, vol. 71, pt. I, Jan. 1952, pp. 26–31.

7. On the Control of Magnetic Amplifiers, R. A. Ramey. *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 2124–28.

8. A New Type of Magnetic Servo Amplifier, W. A. Geyger. *AIEE Transactions*, vol. 71, pt. I, Sept. 1952, pp. 272–80.

9. An Improved Magnetic Servo Amplifier, C. W. Lufcy, A. E. Schmid, P. W. Barnhart. *AIEE Transactions*, vol. 71, pt. I, Sept. 1952, pp. 281–89.

◆

No Discussion

A Magnetic Amplifier for Synchros

R. L. Van ALLEN
ASSOCIATE MEMBER AIEE

Basic Single Core Magnetic Amplifier

THE SYNCHRO magnetic amplifier was developed at the Naval Research Laboratory (NRL)* to eliminate some of the difficulties and failures of the electronic synchro amplifier. At the same time, it provides a demonstration of a system utilizing magnetic circuits of a type which was developed at the Laboratory.[1,2,3]

Since the system draws directly upon a circuit developed within the past few years, it will be worth while to review the principles, first, as applied to proportional amplifiers, and second, as used to sense an error or difference between isolated signals.

To simplify the discussion, let us assume a core material possessing an ideal rectangular hysteresis loop and a circuit with ideal rectifiers and windings with negligible resistance. A basic single-core half-wave magnetic amplifier is shown in Fig. 1. For purposes of this illustration, the power and control windings have a unity turns ratio and are wound in the sense indicated by the dots adjacent to the windings. Voltages E_{ac} and E_{ac}' are sinusoidal, of equal magnitude, and have instantaneous polarity as shown. The magnitudes of the supply voltages E_{ac} and E_{ac}' are such that if either were applied directly to one winding on the core, the time integral of this voltage would cause the core flux to change from knee to knee on the hysteresis loop. The control voltage E_c is rectified alternating current of the same frequency and phase as E_{ac}'. When the polarities are as shown in Fig. 1(A), voltage E_{ac} will appear across the power winding and load in series, while a back voltage is across the control rectifier and I_c is zero. For these voltage polarities the flux will change in the direction marked FIRE.

When polarities of the supply voltage are as shown in Fig. 1(B), the current path is restricted to the control winding, and flux changes on the hysteresis loop will be in a direction marked RESET. This mode of operation resets the flux of the core to a value determined by the time integral of the voltage $E_{ac}'-E_c$. The magnitude of the control voltage E_c will determine the amount of reset and establish the flux of the core at the end of the RESET half-cycle, as shown in Fig. 2(A). During this half cycle, I_c will be the magnetizing current of the core, and I_L will be zero because of the power-circuit rectifiers. During the next half cycle, designated FIRE, voltage E_{ac} will at first appear across the power winding and produce a flux change in the core proportional to the integral of this voltage. As the core flux proceeds to saturation, the current in the load will be only magnetizing current until the instant of time T_x. At the time T_x, the core flux reaches saturation and voltage E_{ac} will then appear across the load, so that I_L suddenly becomes equal to E_{ac} divided by the load impedance for the duration of this half-cycle. Since the change in core flux is proportional to the integral of the voltage across the windings, then during the RESET half-cycle the voltage integral producing a flux change will be

$$\int_0^{T/2} E_{ac}'dt - \int_0^{T/2} E_c dt$$

For the next, or FIRE half-cycle, the voltage integral producing a flux change in the reverse direction will be

$$\int_{T/2}^{T_x} E_{ac}dt = \int_{T/2}^{T} E_{ac}dt - \int_{T_x}^{T} E_{ac}dt$$

Since the total change in flux was the same for both the RESET and FIRE half-cycles, where $E_{ac}'=E_{ac}$, and there is 1 to 1 turns ratio, it follows that

$$\frac{1}{T/2}\int_0^{T/2} E_c dt = \frac{1}{T/2}\int_{T_x}^{T} E_{ac}dt \qquad (1)$$

when averaged over a half-cycle. Thus, voltage is transferred to the load proportional to the applied control voltage and, in the case of a 1 to 1 turns ratio on the core, the average control voltage equals the average load voltage.

Fig. 2(B) shows what happens when the output of the amplifier is minimum, or $E_c=0$. It is true that the load current I_L is not zero during the FIRE half-cycle, because magnetizing current must exist while the flux is changing. However, this minimum load current in the practical amplifier is usually negligible, being several orders of magnitude less than full-load current.

RESPONSE TIME

The integral of the voltage appearing to the load on the FIRE half-cycle depends upon the integral of the signal voltage during the previous half-cycle. So, in order to have steady-state output on the FIRE half-cycle, it is necessary to have steady-state input on the previous RESET half-cycle. Therefore, the response time to reach steady-state output will be the time from initiation of the signal, to the end of the next complete RESET half-cycle. This means the minimum response time will be 1/2 cycle when E_c is applied at the beginning of a RESET half-cycle. Also, the maximum response time will always be less tha $1^1/_2$ cycles, reaching a maximum when E_c is applied immediately after a RESET half-cycle has begun. It should be noted that this response time refers only to steady-state output. Actually, when E_c is applied during a RESET half-cycle, a portion of steady-state output will be obtained on the next FIRE half-cycle, an amount depending upon what percentage of steady-state input was reached during this first RESET half-cycle.

---

* Invented by Dr. Robert A. Ramey, formerly of the Naval Research Laboratory, Magnetic Amplifier Section, now associated with the Westinghouse Electric Corporation.

Paper 53-339, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Pacific General Meeting, Vancouver, B. C., Canada, September 1–4, 1953. Manuscript submitted June 1, 1953; made available for printing July 15, 1953.

R. L. Van Allen is with the Naval Research Laboratory, Washington, D. C.

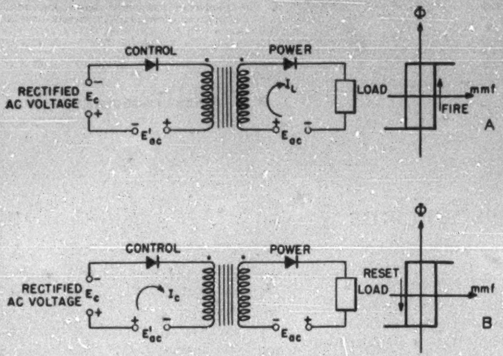


Fig. 1. Polarity of voltages and direction of flux change of single-core magnetic amplifier

A. During FIRE half-cycle

B. During RESET half-cycle

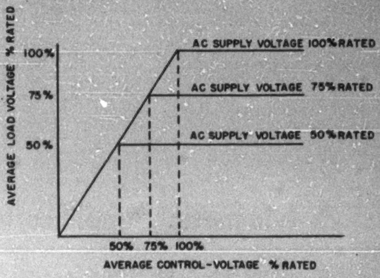


Fig. 3. Transfer characteristics of magnetic amplifier for different supply voltages

Control Element

In addition to the fast time of response of this magnetic amplifier, there are other characteristics which make this circuit especially adaptable for control purposes. For example, the controlling element may be a source of d-c voltage, or a variable resistance, as well as a rectified a-c voltage as shown in the preceding example. The basic requirement of any of these control elements is that they will pass magnetizing current to permit resetting action in the control circuit. Since the controlling element provides an electromotive force which bucks the supply voltage E_{ac}', the output power is proportional to the control voltage. This is a new concept in the field of magnetic amplifiers, which for years have been considered primarily as a current-amplifying device.

Transfer Characteristics

The transfer characteristic (or plot of output versus input) for conventional magnetic amplifiers is a plot of average output current versus average input or control current. In contrast, the NRL-developed magnetic amplifier, having as the control element a source of electromotive force, establishes a linear transfer relationship between average control voltage and average load voltage, see Fig. 3. This linear relationship holds for a range of control voltages from $E_c=0$ to $E_c=E_{ac}'$, providing rectifier leakage is not excessive. It is very significant, in regard to the synchro amplifier, that, provided the ratio E_{ac}/E_{ac}' remains constant, this linear transfer characteristic is not affected through the control range by changes in the supply voltage. Of course, the maximum power-handling capacity for a given load of the amplifier is reduced as the supply voltages are reduced, as shown in Fig. 3.

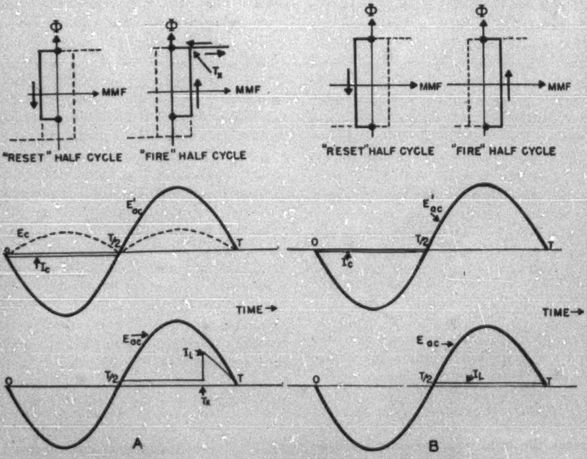


Fig. 2 Voltage and current wave forms and flux change of single-core magnetic amplifier

A. With signal voltage

B. Without signal voltage

Special Circuit Characteristics

It is shown in Fig. 3 and by equation 1 that the average voltage applied to the load is proportional to the average control voltage E_c. This voltage output to the load is a direct result of E_c opposing E_{ac}'. Similarly, in the absence of E_c, a reduction in E_{ac}' will produce an output voltage, $\int V_{out}dt=\int(E_{ac}-E_{ac}')dt$. This characteristic of the circuit to transfer to the load a voltage proportional to the difference between the supply voltages E_{ac} and E_{ac}' is unique and bilateral. To illustrate this bilateral action, assume a single-core circuit having the parameters indicated in Fig. 4. When E_{ac} is greater than E_{ac}', circuit no. 2 fires, owing to saturation of the core at one end of the hysteresis loop, and a voltage proportional to $E=E_{ac}-E_{ac}'$ is applied to R. Conversely, when E_{ac}' is greater than E_{ac}, circuit no. 1 fires, owing to saturation of the core at the opposite end of the hysteresis loop, and the voltage across R' is proportional to $E=E_{ac}'-E_{ac}$. This simple circuit, in addition to exhibiting a bilateral voltage transfer proportional to the difference between E_{ac} and E_{ac}'

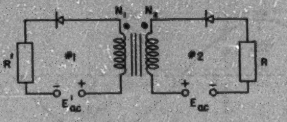


Fig. 4. Single-core bilateral network
Square-loop core material
$N_1 = N_2$ $R = R'$ $E_{ac} = E_{ac}'$

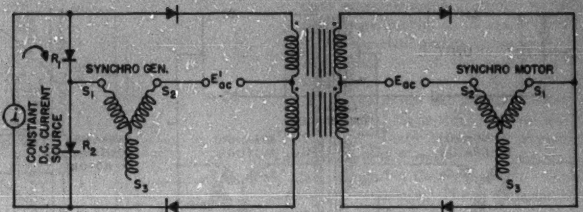


Fig. 6. Full-wave synchro magnetic amplifier

also possesses another unique property of error sensing. Error sensing is due to the high remanence flux of the square-loop material, which exhibits a memory action, so that E_{ac} in circuit no. 2 will act upon R during one half-cycle, depending upon how E_{ac}' in circuit no. 1 acted upon the core in the preceding half-cycle. Conversely, E_{ac}' will act upon R' in accordance with the action of E_{ac} upon the core in the preceding half-cycle. Thus, an output is obtained as a result of error sensing or a comparison that is made between E_{ac} and E_{ac}', with an inherent delay of 1/2 cycle introduced.

Application to Synchro Amplifier

The application of magnetic amplifiers to function as a synchro amplifier may be accomplished in several ways. However, in order to have accuracy, fast response, high sensitivity, and stability, one circuit was chosen from several alternatives. This circuit is a multiple application of the basic Ramey magnetic amplifiers.

Fig. 5 is a block diagram of the complete system showing full-wave magnetic amplifiers. In the block diagram, it is indicated how the control voltages for the six units are obtained from the stator of a synchro generator, with each pair of stator terminals controlling two magnetic amplifiers. Thus:

1. Voltage S_2S_3 controls magnetic amplifiers A and A'
2. Voltage S_1S_3 controls magnetic amplifiers B and B'
3. Voltage S_1S_2 controls magnetic amplifiers C and C'

The output of each pair of amplifiers supplies the respective synchro motor terminals. Thus:

1. Amplifiers A and A' supply synchro motor terminals S_2S_3
2. Amplifiers B and B' supply synchro motor terminals S_1S_3
3. Amplifiers C and C' supply synchro motor terminals S_1S_2

The apparent duplication of amplifier units for each input and output serves a specific purpose, which will be explained later.

Each of the amplifiers A, A', B, B', C, and C' is an identical full-wave amplifier, as shown in Fig. 6. A constant current source i and two additional rectifiers, R_1 and R_2, are in the control circuit, see Fig. 6, to insure that the control current will not rise above the magnetizing current of the core. As was previously shown, it is necessary to permit magnetizing current in the control winding only during the RESET half-cycle. The output load on the synchro amplifier unit, see Fig. 6, consists of the winding between a pair of terminals on the synchro motor bus, in which a voltage is induced varying in magnitude and polarity with the position of the rotor. This presents a problem in magnetic-amplifier loading, since the impedance of the load varies with the number of synchro repeaters used, and the voltage induced in the load varies with the angular position of these repeaters.

When a synchro shaft is rotated through 360 degrees and the rotor is energized with a single-phase voltage, the stator terminal voltages will appear as shown in Fig. 7. Thus, during one complete revolution the stator terminal voltage (for example $V_{S_1S_3}$) will change in magnitude and polarity, having the same polarity as the rotor voltage for 180 degrees rotation and reversing polarity for the other 180 degrees. Since the magnetic amplifier is to respond to the difference between motor and generator terminal voltage, it must also respond to a change in polarity of this voltage. To provide for polarity reversal, the magnetic-amplifier circuit shown in Fig. 8 is used.

In essence, this circuit has two full-wave Ramey magnetic amplifiers connected for parallel operation, except that E_{ac} and E_{ac}' of one unit are in reverse polarity with E_{ac} and E_{ac}' of the other unit. This is arranged by supplying E_{ac} and E_{ac}' from two center-tapped windings of a transformer. By making $E_{ac} = E_{ac}'$ and equal to or greater than the maximum synchro terminal voltage, the direction of current through the synchro stator windings will be determined by the instantaneous polarity of voltages E_{ac} and E_{ac}'.

Since it is the voltage available to the core that we are concerned with in both

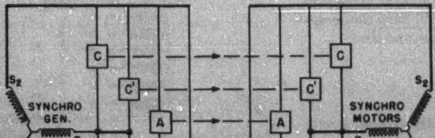


Fig. 5. Block diagram of synchro magnetic amplifiers

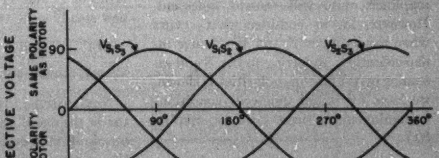


Fig. 7. Synchro stator voltage as a function of angular position of the rotor

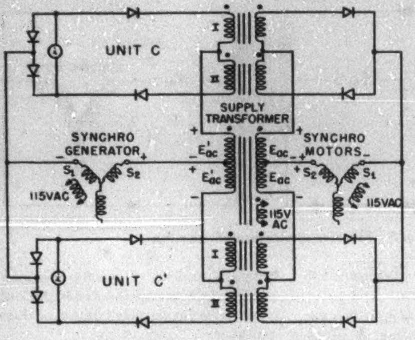


Fig. 8. Two full-wave magnetic amplifiers arranged to accommodate change in polarity of synchro stator voltage

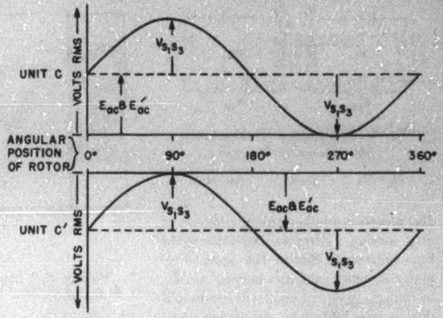


Fig. 9. Voltage available to the synchro magnetic-amplifier core as a function of the angular position of the rotor

the power and the control circuit, let us examine Fig. 9. In this figure, root-mean-square voltage is plotted versus rotor displacement. It should be noted that this a volt-displacement plot rather than a volt-time plot. The voltage available to the core is the algebraic sum of the supply voltage and the synchro terminal voltage. This is equivalent to having a single supply voltage $V_t = E_{ac} + V_{S_1S_2}$ (motor), in the power circuit, and $V_t' = E_{ac}' + V_{S_1S_2}$ (generator), in the control circuit. As long as the ratio of V_t/V_t' remains a constant equal to the turns ratio of the windings, the magnetic amplifiers are quiescent, which is equivalent to E_{ac}/E_{ac}' remaining constant and with no signal voltage, in the basic circuit shown in Fig. 1. Also, in Fig. 3, the values of E_{ac} and E_{ac}' may change, but as long as the ratio E_{ac}/E_{ac}' remains constant, the magnetic amplifier operates along the same linear transfer characteristic. This is the condition that exists in the synchro amplifier as the motor and generator assume the same angular position through a complete revolution of 360 degrees. If the angular position of both are the same, V_t and V_t' will change proportionally and all magnetic-amplifier units will remain quiescent. However, let us consider what occurs when the voltages V_t and V_t' change independently. Referring to Fig. 8 and assuming instantaneous polarities as shown, we have for initial quiescent conditions (assuming unity turns ratio for simplicity):

Unit C

$V_t = V_t'$
$V_t = E_{ac} + V_{S_1S_2}$ (motor)
$V_t' = E_{ac}' + V_{S_1S_2}$ (generator)

and

Unit C'

$V_t = V_t'$
$V_t = -E_{ac} + V_{S_1S_2}$ (motor)
$V_t' = -E_{ac} + V_{S_1S_2}$ (generator)

Now assuming the synchro generator is displaced so as to increase its terminal voltage, $V_{S_1S_2}$ generator, we have

Unit C	Unit C'
$V_t' > V_t$	$V_t' < V_t$

Neither core of unit *C* fires because of blocking by the constant current source and rectifiers. Core I of unit *C'* will reset a smaller amount, so that on the next half cycle an excess voltage equal to the difference between V_t' and V_t will appear across the motor synchro terminals S_1S_2. Since the amplifiers are full-wave, core II of unit *C'* will fire on the second half-cycle following initiation of the signal. The current through the synchro motors will be in such a direction as to produce torque to position the rotors and to restore quiescence again. As soon as the motor rotor is positioned, the magnitudes of V_t and V_t' will have changed to new steady-state values and then

Unit C	Unit C'
$V_t' = V_t$	$V_t' = V_t$

Next, assume that the synchro generator is displaced such as to decrease its terminal voltage, $V_{S_1S_2}$ generator; for the same instantaneous polarities, see Fig. 8, we have

Unit C	Unit C'
$V_t' < V_t$	$V_t' > V_t$

Core II of unit *C* will reset a smaller amount, and this core will fire on the next half-cycle. Similarly, core I of unit *C* will fire on the second half-cycle and torque in the synchro motors will be in a direction to restore quiescence, until again

Unit C	Unit C'
$V_t' = V_t$	$V_t' = V_t$

It is evident that the two units *C* and *C'* act to provide torque in opposite directions irrespective of the polarity of the synchro's terminal voltage, with the direction of torque dependent on the polarity of the synchro terminal voltage. In other words, for one half revolution of the generator rotor relative to the motor rotor, unit *C* will provide clockwise torque, and for the other half, counterclockwise.

When there is a displacement of the motor shaft with the generator in a locked position, a torque is produced resisting this displacement. The action of a core firing is again a function of the voltages V_t and V_t' and, since a displacement of

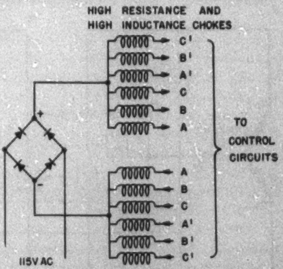


Fig. 10. Diagram of constant-current d-c supply

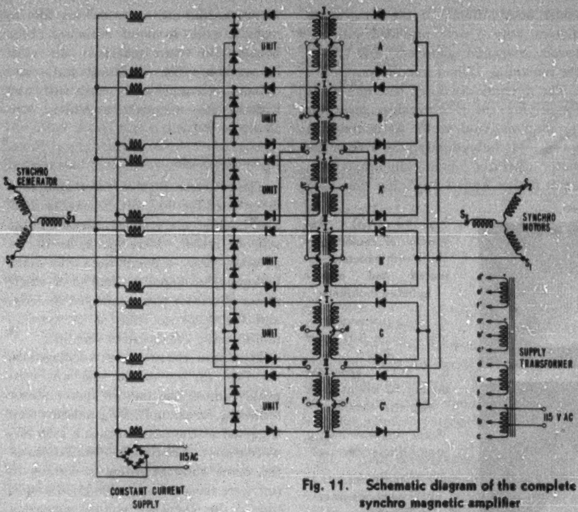


Fig. 11. Schematic diagram of the complete synchro magnetic amplifier

the motor rotor causes a change in the synchro motor terminal voltage, there is a consequent unbalance produced by a change in V_t instead of V_t'.

In going through the step-by-step analysis of the operation of the synchro amplifier, the main purpose of the constant-current sources becomes readily apparent—that of blocking other than magnetizing currents in the control circuit. For the case of a generator-initiated signal in the circuit, Fig. 8 the voltage V_t' is seen to increase in one unit and at the same time to decrease in the other. Dependent on past history and upon the operating point along the hysteresis loop, an increase in V_t' may cause saturation on the reset half-cycle. This tends to produce large currents in the control circuit, which are prevented by the constant-current source and associated blocking rectifiers.

All of the constant-current sources are supplied from a single full-wave bridge rectifier using the 115-volt a-c supply, Fig. 10. In both the positive and negative leads, high-impedance chokes are used, so that the d-c current is maintained nearly constant because of the high-d-c resistance of the chokes, and a-c isolation between units is assured by the large inductance of the chokes.

Complete System

Fig. 11 is a schematic diagram of the complete synchro amplifier. For the sake of clarity, the power and control circuits are diagrammed separately, since the two circuits are electrically isolated except for inductive coupling through the cores. This circuit is a composite of three magnetic amplifier circuits of the type shown in Fig. 8. A simplification of the complete system is obtained by eliminating two of the center-tapped transformer windings. Here, Units *A*, *A'*, *B*, and *B'* are seen to share common a-c supplies, E_{ao} and E_{ao}', whereas units *C* and *C'* require different a-c supplies. Therefore, the whole system uses eight a-c supplies, in the form of four center-tapped windings. These four center-tapped windings are provided from a single transformer, with its primary supplied from the 115-volt line.

Results of Tests

In the investigation of this magnetic amplifier, an attempt was made to determine the accuracy, time of response, and gain. Measurements were made with satisfactory precision on static and dynamic accuracy and time of response. No satisfactory measurement of gain was obtained because of practical problems involved.

ACCURACY

The static and dynamic accuracy of the synchros was tested both with and without the magnetic amplifier. As a basis for comparison, both static and dynamic accuracy checks were made on a

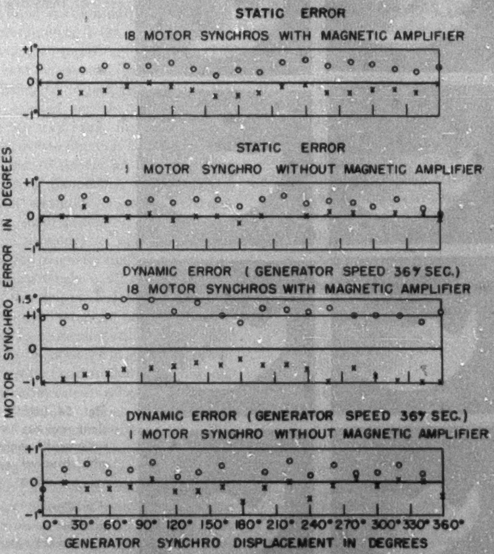


Fig. 12. Static and dynamic accuracy of the system

○—Rotation clockwise
×—Rotation counterclockwise

Bureau of Ordnance *5-G* synchro generator driving a single *5-F* synchro motor, direct-coupled without magnetic amplifiers. Both of these units were matched synchros selected to have an accuracy of ±0.5 degrees. This same motor-generator pair was tested for accuracy with the magnetic amplifiers in service, and with 17 additional synchro motors of the same size paralleling the *5-F* synchro motor under observation. Of the 18 synchro motors, only 6 were provided with external magnetic dampers and dials; the remaining 12 had free shafts.

The dynamic accuracy test was made by driving the *5-G* synchro generator from an external motor at a constant speed, and photographing the dials of the motor-generator pair simultaneously. Light flashes from a Strobolux were used to make exposures on a moving film to make a point-by-point accuracy check between the motor-generator pair while in rotation. The results of static and dynamic accuracy tests, both with and without the magnetic amplifiers, are shown in Fig. 12.

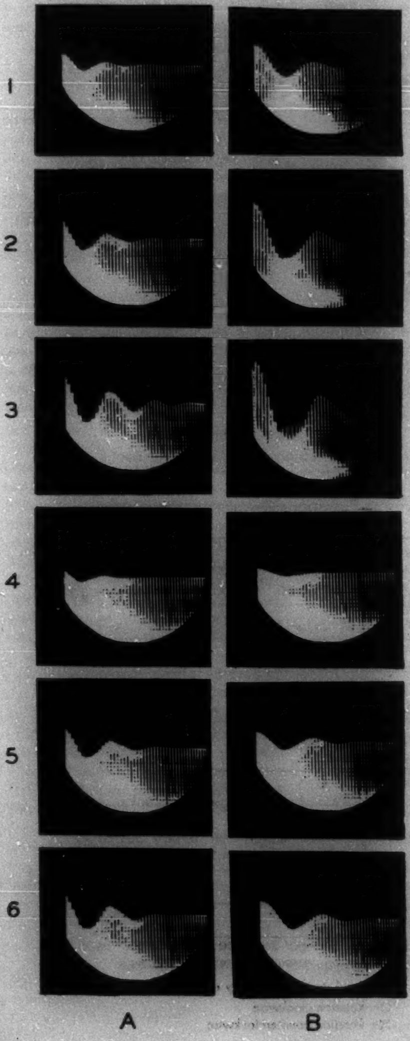


Fig. 13 (left). Response of motor synchros with generator-initiated and motor-initiated transients

A. 5-G generator driving a single 5-F motor direct

B. 5-G generator driving 18 motors with magnetic amplifiers

Motor response from a motor displacement of

1. Ten degrees
2. Twenty degrees
3. Thirty degrees

Motor response from a generator displacement of

4. Ten degrees
5. Twenty degrees
6. Thirty degrees

RESPONSE TIME

The response time of the system was determined for step displacements initiated in either the synchro generator or the synchro motor. Here again, as in the accuracy tests, a comparison was made between the response time of a single motor-generator pair connected directly, and the response time of 18 synchro motors using the magnetic amplifier.

To obtain the transient oscillograms, see Fig. 13, an oscilloscope was connected across one pair of motor stator terminals, for example, S_2S_3, and the sweep triggered at the same instant a step displacement was initiated. Oscillations of the motor rotor or rotors in coming to rest were recorded as a change in amplitude of the 60-cps a-c voltage appearing at the stator terminals of the synchro motor. By choosing the proper rotor position to initiate the transients, the 60-cps voltage $V_{S_2S_3}$ was amplitude-modulated to be very nearly proportional to displacement; the 60-cps voltage peaks provided a convenient time reference.

Generator-initiated transients were produced by switching from one to the other of two locked generators having appropriate angular displacement. Motor-initiated transients were made by displacing a single motor rotor and triggering the sweep at the same instant the displaced rotor was released.

The response time of the amplifiers alone, without the influence of the natural period of the synchro motors, was obtained as follows. The synchro motors were positioned by a generator (no. 1) and then all motor shafts were locked. The control was then switched to another generator (no. 2), displaced 20 degrees,

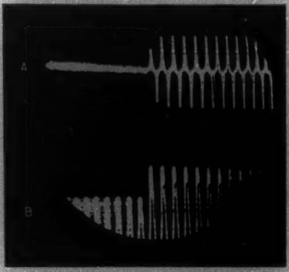


Fig. 14 (right). Transient response of magnetic amplifiers obtained with blocked synchro rotors

A. Current supplied to motor synchros with blocked rotors

B. Transient introduced in generator synchro voltage and the oscilloscope sweeps triggered a short time before the instant of switching. One *y*-axis of a dual-beam oscilloscope was connected across a small resistance in one synchro-motor lead, so that motor current was recorded as a function of time. The other *y*-axis was connected across a pair of control circuit terminals to record the time that the generators were switched.

Fig. 14 shows the result of this measurement, and it may be noted that the response time is essentially one cycle. This response time correlates with the dynamic accuracy tests made on the system, which show an average angular lag of the motor synchros of approximately 3/4 degree for each direction of rotation, see Fig. 12. Since the speed of rotation for this test was 36 degrees per second, then 3/4 degree corresponds to a time lag of 0.02 second which, in terms of 60-cycle power-supply frequency, is a response time of 1.2 cycles. When the speed of rotation is reduced one half, or 18 rpm, the dynamic error is correspondingly reduced by one half.

GAIN

One of the greatest advantages of this particular magnetic amplifier is that gain and time of response are not interrelated. Furthermore, the output current is not necessarily a function of input current, as it is for conventional amplifiers; instead, the gain of this circuit is more of a function of the ratio of input to output circuit impedances. Of course there are practical limitations to this type of magnetic-amplifier circuit, such as winding resistance, a-c power-supply impedance, rectifier back-leakage, and the fact that even the best core materials do not have their remanent flux density equal to their saturation flux density. In spite of these practical limitations, the results obtained with the synchro magnetic amplifiers are good.

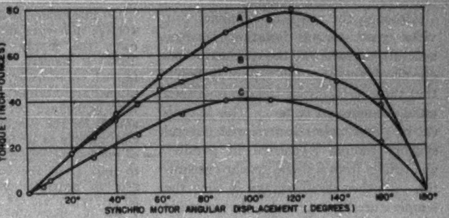


Fig. 15. Static output torque with and without magnetic amplifiers

A. Eighteen motor synchros with magnetic amplifier

B. Two motor synchros with magnetic amplifiers

C. Two motor synchros without magnetic amplifiers

Table I. Design Data

Item	Specifications
Magnetic-amplifier cores	29-gauge Hypersil toroidal strip wound 2⁷/₈-inch outside diameter; 2¹/₁₆-inch inside diameter; 1-inch height
Magnetic-amplifier windings	2,000 turns no. 22 (power circuit) 2,000 turns No. 29 (control circuit)
Supply transformer	Primary: 115 volt Secondaries: two 180-volt center-tapped windings for E_{ac} supplies, two 170-volt center-tapped windings for E_{ac}' supplies
Rectifiers	Power: each rectifier shown schematically in the power circuits is made up of two *G-10* power germanium rectifiers in parallel Control: all rectifiers are *1N39* germanium diodes Constant current source: (Fig. 10) four 100-milliampere 120 volt selenium rectifiers
Chokes	Constant current source: (Fig. 10) twelve 5,500 ohm, 250-henry chokes

To obtain measurements of the gain of the complete synchro magnetic amplifier under operating conditions posed a problem. Since the output of the amplifier is, in the final analysis, intended to provide torque through the respective motor synchros, it was believed that gain measurements could be obtained from torque measurements. Although output torque could be readily measured, the input torque was essentially zero, because of current limiting in the control circuit of the synchro amplifier. This limiting action kept the generator stator current to 11 milliamperes or less, regardless of the torque exerted on the motor shafts. This is not surprising, since the system is not designed to function as a torque amplifier, but to deliver an output torque based on a comparison of motor and generator terminal voltages. Nevertheless the small generator stator current which is necessary as magnetizing current for the magnetic amplifiers does produce an extremely small reacting torque (which was impractical to measure, because approximately 10 milliamperes are required to overcome static friction alone). Owing to the presence of static friction in the generator synchro shaft, which would swamp out measurement of input torque, overly optimistic figures of gain would result from torque measurements. Fig. 15 gives the results of static torque measurements and, for purposes of comparison, the torque displacement curve of a single *5G* generator synchro, supplying two *5F* motor synchros, is included. It can be seen from this graph that the maximum output torque obtained from a motor synchro, with the magnetic amplifier supplying 18 motor synchros, is twice that obtained when two motors and one generator are connected directly, and also that there is an increase in the maximum torque available when more motor synchros are added in parallel, as a load on the magnetic amplifiers. This results from one characteristic of this magnetic-amplifier circuit, which performs more efficiently when coupled to a low-impedance load (in this case, a number of synchro-motor impedances in parallel).

Conclusions

The magnetic amplifier described in this paper provides power amplification, so that the signal obtained from a single synchro generator may be used to position a large number of synchro motors connected in parallel.

This particular magnetic synchro amplifier was not designed as a replacement for any specific electronic synchro amplifier, but rather to demonstrate a principle. However, the size and weight of this particular magnetic amplifier and its performance indicate that a system designed to replace certain electronic amplifier systems would offer considerable saving in weight and space. The minimum size and weight of a magnetic synchro amplifier of this type designed for any particular installation will depend upon:

1. Type and number of synchro motors;
2. Mechanical load on the synchro motors;
3. Static and dynamic accuracy requirement;
4. Speed of response desired.

Table II. Space and Weight

Item	Weight
Twelve magnetic-amplifier cores with windings at 2¹/₄ pounds each	27 pounds
Supply transformer	11 pounds
Twelve chokes, at 1¹/₂ pounds each	18 pounds
Rectifiers	4 pounds
Total weight (excluding chassis and hardware)	60 pounds

Entire system may be conveniently housed in a volume of 2¹/₂ cubic feet (approximately 12 by 10 by 30 inches).

Although the magnetic amplifier provides good torque isolation between synchro generator and the synchro motors one disadvantage of this type of system is that there is no torque isolation between parallel-connected motors. This is typical of any synchro system with several synchro motors connected in parallel, and driven from a single synchro generator. Another disadvantage is that the temperature limitations on the magnetic amplifier are dictated by the germanium rectifiers, which have a restricted operating-temperature range.

References

1. ON THE MECHANICS OF MAGNETIC AMPLIFIER OPERATION, **Robert A. Ramey.** *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 1214–23.

2. ON THE CONTROL OF MAGNETIC AMPLIFIERS, **R. A. Ramey.** *AIEE Transactions*, vol. 70, pt. II, 1951, pp. 2124–28.

3. THE SINGLE-CORE MAGNETIC AMPLIFIER AS A COMPUTER ELEMENT, **R. A. Ramey.** *AIEE Transactions*, vol. 71, pt. I, Jan. 1953, pp. 442–46.

◆

Discussion

L. J. Johnson (D and R, Ltd., Santa Barbara, Calif.) and **Dr. S. E. Rauch** (Associate Professor, University of California, Los Angeles, Calif.): The author has presented details in his discussion of the synchro system which lead the reader to expect equivalent center-tap power transformers in both the control and load side of the amplifiers. For example, in Table I it is shown that the number of turns in the power and control circuits are equal. In addition, it is necessary that the control voltage equal the output voltage, since this condition would have to exist if the synchro generator and synchro motors were used without the magnetic amplifier. However, Table I shows a 180-volt center-tap winding for E_{ac} supplies, and a 170-volt center-tap winding for E_{ac}' supplies. This 10-volt difference between the power- and control-circuit voltages has not been explained in the paper. Since the voltage difference is a vital requirement in obtaining the stability of the synchro system, it is of general interest to have this phenomenon discussed in detail.

Let us first consider the bilateral amplifier shown in Fig. 4. It can be seen that if $E_{ac}'>E_{ac}$, then circuit no. 1 will fire and circuit no. 2 will not fire. Thus, the magnetic flux will traverse an asymmetrical hysteresis loop. The major portion of this loop will be in the flux direction determined by the magnetomotive force produced by the currents in circuit no. 1. Now suppose that a reversal of voltage difference occurs, such that $E_{ac}>E_{ac}'$. Then the minor loop must progress towards the opposite end of the major loop. The rate of progress is determined only by the voltage difference $E_{ac}-E_{ac}'$. Thus, the time to reach steady state is a function of the difference voltage. It is possible for the time delay to be of the order of magnitude of a second or more to reach steady state.

Let us examine in detail the effects of the voltage difference in the circuit of Fig. 8, using the supply voltages as listed in Table I (namely, $E_{ac}'=170$ volt center-tapped transformer, $E_{ac}=180$-volt center-tapped transformer). Assume the magnitude of voltage between the terminals S_1 and S_2 for the synchro generator and synchro motors to be 50 volts, with polarities as marked in Fig. 8. The voltage applied to unit C in the control circuit will be synchro generator voltage, plus E_{ac}', or $50+85=135$ volts. The voltage to unit C on the load side is equal to E_{ac} plus synchro motor voltage, namely $90+50=140$ volts. It is to be noted that equal turns exist in the control and power windings. Since there is a 5-volt difference in favor of the load circuit, and the amplifier employs the Ramey principle, the 5 volts must be absorbed while the amplifier is firing.

Now consider the control and load circuits which contain unit C'. The voltage to the control circuit equals E_{ac}' minus synchro generator voltage, or $85-50=35$ volts. Similarly the voltage to the load circuit is E_{ac} minus synchro motor voltage, or $90-50=40$ volts. Again, there exists a 5-volt excess in load circuit which is absorbed when unit C' is firing. From these observations it follows that unit C and unit C' fire simultaneously with equal magnitudes when the synchro motors and synchro generator are in perfect alignment. As the circuit in Fig. 8 is in triplicate for the complete synchro-amplifier circuit, it must be concluded that all amplifiers fire simultaneously with 5-volt magnitudes when the synchro generator and synchro motors are in perfect alignment. This phenomenon ensures that all cores work around minor loops which are located at the firing end of the major loop. As a result, there is no time delay involved in affecting a change of flux from the reset end to the firing end of the major loop when a transient error is encountered. This would not necessarily be the case, if $E_{ac}=E_{ac}'$. Thus, the significance of the voltage difference has been established. Furthermore, stability of the circuit has been improved by eliminating the possibility of pure time delays which arise when the flux must drift from the reset to the firing end of the major loop. It is well known in the servo field that pure time delays are very detrimental to the stability of systems.

The designer will immediately raise the question concerning the process in which the 5-volt differential was established for the specific synchro system reported upon by the author. In general, the required magnitude of the voltage differential is empirically obtained and is dependent upon core material characteristics. The permeability in the saturated region must be of sufficient magnitude to limit peak current values when all amplifiers are firing in equal magnitudes; namely, during the steady-state quiescent condition. During this period the firing currents do not necessarily flow through the synchro motors, but may flow directly from one amplifier unit to the other. For example, in Fig. 8 currents may flow out of unit C into unit C' and through the entire supply transformer winding. It may be noted further that the resistance is very low, and thus a small reactance exhibited by the windings on the saturated core will significantly limit the peak values of these currents. As the required voltage differential is always proportionally small in comparison to supply transformer voltages, it is only necessary that the core characteristic exhibit sufficient permeability magnitude for current limitation near the knee for small firing angles. It is highly desirable for the permeability beyond the knee to rapidly decrease to its minimum value for larger firing angles.

Hypersil provides characteristics which appear to be the most desirable for the application reported in this paper. Deltamax is less desirable, since its permeability change at the knee is too abrupt and would necessitate a much closer control of the voltage difference herein discussed. A more detailed analysis of the effects of core materials on circuits employing magnetic amplifiers can be found in the references.

REFERENCES

1. THE EFFECT OF CORE MATERIALS ON MAGNETIC AMPLIFIER CIRCUITS, **Leo J. Johnson.** *AIEE Transactions*, vol. 71, pt. I, Jan. 1952, pp. 26–31.

2. SERIES-CONNECTED MAGNETIC AMPLIFIER WITH INDUCTIVE LOADING, **Thomas G. Wilson.** *AIEE Transactions*, vol. 71, pt. I, Jan. 1952, pp. 101–10.

3. HIGH SPEED MAGNETIC AMPLIFIERS, **Leo J. Johnson.** *Electrical Manufacturing*, New York, N. Y., Nov. 1952, pp. 98–101.

R. L. Van Allen: The author wishes to thank Mr. Johnson and Dr. Rauch for an interesting discussion. However, it should be stated that this article was submitted as a technical paper to demonstrate the workability of a principle set forth by Dr. Ramey. There are many design problems which may arise in adapting this principle to a specific synchro system. This paper was not intended as a guide to the designer, but since several questions have been raised regarding design, the author would like to add to the discussion some notes obtained from his experience with the system.

Early work with this magnetic synchro amplifier was performed on a model which incorporated the use of selenium rectifiers in the load circuits, and although the selenium rectifiers were the best available, they adversely affected the performance of the system. Using selenium rectifiers, it was found to be necessary to establish a quiescent firing angle by making the load power supply E_{ac} much larger than the control-circuit power supply, E_{ac}'. In addition to preventing pure time delays, this quiescent firing angle also served an additional function of biasing all magnetic amplifiers up on the transfer characteristic curve. The ideal transfer characteristic, see Fig. 3, does not truly represent the practical amplifier transfer characteristic. Because of imperfect core materials and forward resistance of rectifiers, the practical transfer characteristic curve exhibits a toe at the base of the curve. For small output currents at small values of control voltage, the slope of the curve is not as steep as the rest of the curve. Therefore, unless some means of biasing is provided, the magnetic synchro amplifier will be insensitive to small angular changes of the generator. By operating in the linear regions of the transfer characteristic, sensitivity, static accuracy, and stability are improved. There is a disadvantage in biasing for a quiescent firing angle;

that is, all cores in each amplifier unit must be carefully matched, or else static accuracy will be impaired. In the paper, it was mentioned that alternate units such as *C* and *C'* provide either clockwise or counterclockwise torque. It is evident that when all units have a quiescent firing angle, all counteracting torques must be equal, or else the motor synchros will seek a position not in alignment with the generator, thus impairing the static accuracy of the system. It was found that most of the quiescent current does pass through the motor windings, and hence must be taken into account.

During the course of the investigation of this system, germanium power rectifiers became available, and when they were used in the power circuits of the magnetic amplifiers, stability of the system could be obtained with E_{ac} nearly equal to E_{ac}' because of an increase in the slope of the transfer characteristic. It was found, however, that for small angle sensitivity, static accuracy, and improved time of response, a quiescent firing angle was still desirable, for these reasons as well as for the elimination of pure delay time. The amount of voltage difference between E_{ac} and E_{ac}' was arrived at empirically, but the absolute value of this difference is not extremely critical and may vary several volts. But, it is best to use as little voltage difference as possible to keep quiescent current low and still maintain good sensitivity and time of response.

The problems of stability and speed of response are closely related, and it was found in the course of the investigation that factors which influenced the speed of response of the individual amplifier are most important in establishing a stable system. If we carefully examine the basic amplifier of Fig. 6, it becomes apparent that it is not inherently a half-cycle response amplifier, as is the more basic circuit invented by Dr. Ramey, see Fig. 1.

Referring to the circuit which is illustrated in Fig. 6, let us assume for initial conditions that the motor and generator are in alignment and their terminal voltages are each 50 volts with polarities aiding E_{ac} and E_{ac}'. In the control circuit, the resetting voltage is $E_{ac}'+50$ volts, and in the load circuit, $E_{ac}+50$ volts. Further, let us assume that instantaneously the generator voltage is increased to 70 volts at the beginning of a cycle in which the top core will fire and the lower core reset. The step increase in generator voltage will cause a voltage $E_{ac}'+70$ to appear across the control windings of the lower core, and transform to the power winding of the lower core $E_{ac}'+70$ volts. Since, in the power circuit, the source voltages are $E_{ac}+50$ volts (and with E_{ac} 5 volts greater than E_{ac}), there is still a net of 15 volts excess in the direction of the load rectifier of the resetting core, so that the lower power rectifier will conduct and reduce the resetting voltage to $E_{ac}+50$ volts. Thus the required amount of reset does not occur in 1/2 cycle, but must take several cycles to reach steady state. If either the synchro motor impedance or the impedance of E_{ac} supply is appreciable, and the current through the firing core causes a drop in the terminal voltage, then there will be a greater excess of voltage transformed by the resetting core, further increasing the time of response. For this particular system, it was found that six type-5*F* synchro motors in parallel reduced the effective impedance of the load for minimum time of response, which is of the order of 1 cycle. When more than six motors are used, no appreciable decrease in response time can be realized, but when less than six motors are connected, the time of response increases, and consequently more overshoots will result from a generator-initiated transient, unless the load impendance is reduced. It is impractical to use this system for fewer than six motors synchros of this size, so this is not considered a serious disadvantage.

In summary, the most important factors which must be considered for good performance of the system are:

1. Low internal impedance of E_{ac} supply and load;
2. Power rectifier should have low forward resistance;
3. Biasing for quiescent firing angle;
4. Careful core matching.

Type-N Carrier Telephone Deviation Regulator

E. H. PERKINS
NONMEMBER AIEE

J. J. MAHONEY
NONMEMBER AIEE

THE type-*N* carrier system[1] was introduced into the field 3 years ago to satisfy the increasing demand for short-haul toll telephone trunks. The system provides flat regulation for the group of 12 channels at each repeater and the receiving terminal. Individual channel regulation is provided only at the receiving terminal. The flat group regulation is adequate for longer systems, but for system lengths greater than about 150 miles, the transmission variation caused by temperature change could exceed the individual channel regulating ability of the receiving terminal. If this were to occur, the channels would not meet the severe standards specified for them as a component of a multilink toll connection.

To prevent this condition, a regulator has been designed which can be inserted into the line and which provides the additional transmission regulation required. The regulator performs the function of re-establishing correct channel levels as would the insertion into the line of two terminals back-to-back, but at a much lower cost. This regulator is activated by all of the transmitted channel carriers. It determines the departure of each channel carrier from its nominal value and regulates the line transmission by thermistor-controlled networks, so as to produce a minimum residual departure from nominal transmission. The correction includes four shapes designated slope, bulge, cubic, and quartic. Various combinations of these shapes very closely approximate the departures experienced in the cable and repeaters. This regulator controls the transmission variations sufficiently well so that temperature changes are no longer the controlling factor in limiting system length.

Transmission Requirements

The transmission variations occurring in a system are proportional to the length of the system, the number of repeaters involved, and the extent that transmission through each of these components varies with operating temperature. Thus, the amount of transmission variation is a function of the temperature range of the area where the system is used. This range varies considerably over the country. Fig. 1 presents data compiled by the United States Weather Bureau to show the normal annual temperature variations in the United States. A near maximum figure, not this normal variation, must be used for circuit design so that service standards will not be deteriorated on more than a few extreme days that exist each year in both winter and summer. An analysis of Weather Bureau data for many cities shows that a factor of three times the normal variation is representative of the near maximum figure desired, and accordingly that value has been considered as a basis for the regulator performance. This establishes, for example, a temperature swing of about 75 degrees

---

Paper 53-334, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Pacific General Meeting, Vancouver, B. C., Canada, September 1–4, 1953. Manuscript submitted May 25, 1953; made available for printing July 13, 1953.

E. H. PERKINS is with Bell Telephone Laboratories, Inc., Murray Hill, N. J., and J. J. MAHONEY is with Bell Telephone Laboratories, Inc., New York, N. Y.

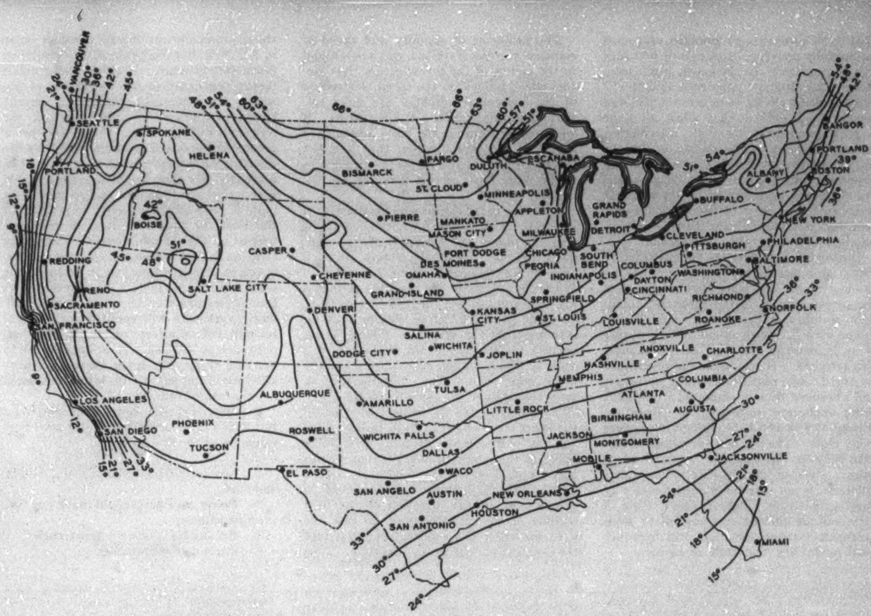

VANCOUVER
SEATTLE
SPOKANE
HELENA
PORTLAND
BOISE
REDDING
RENO
SACRAMENTO
SAN FRANCISCO
SALT LAKE CITY
LOS ANGELES
SAN DIEGO
PHOENIX
TUCSON
CASPER
CHEYENNE
DENVER
ALBUQUERQUE
ROSWELL
EL PASO
BISMARCK
FARGO
ST. CLOUD
PIERRE
DULUTH
MINNEAPOLIS
MANKATO
MASON CITY
FORT DODGE
DES MOINES
OMAHA
GRAND ISLAND
KANSAS CITY
SALINA
WICHITA
DODGE CITY
JOPLIN
TULSA
AMARILLO
WICHITA FALLS
DALLAS
WACO
SAN ANGELO
AUSTIN
SAN ANTONIO
HOUSTON
NEW ORLEANS
ESCANABA
APPLETON
MILWAUKEE
CHICAGO
PEORIA
SOUTH BEND
GRAND RAPIDS
DETROIT
INDIANAPOLIS
SPRINGFIELD
ST. LOUIS
LOUISVILLE
COLUMBUS
DAYTON
CINCINNATI
CLEVELAND
PITTSBURGH
BUFFALO
ALBANY
BANGOR
PORTLAND
BOSTON
NEW YORK
PHILADELPHIA
BALTIMORE
WASHINGTON
RICHMOND
NORFOLK
ROANOKE
KNOXVILLE
CHARLOTTE
NASHVILLE
MEMPHIS
LITTLE ROCK
COLUMBIA
ATLANTA
AUGUSTA
BIRMINGHAM
JACKSON
MONTGOMERY
MOBILE
JACKSONVILLE
MIAMI
12° 24° 30° 36° 42° 45°
48° 51° 54° 57° 60° 63° 66°
18° 15° 12° 9° 6°
42° 51° 45° 48°
66° 63° 60° 57° 51°
51° 54° 39° 36°
45° 42° 39°
36° 33° 30°
30° 27° 24°
21° 18° 15°
12° 15° 21° 27° 33°
33° 30° 27° 24°
24° 21° 18° 15°

Fig. 1. Normal annual temperature variation in United States, compiled by the United States Weather Bureau

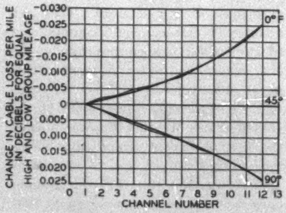


Fig. 2. Transmission loss variation with temperature in 19-gauge cable

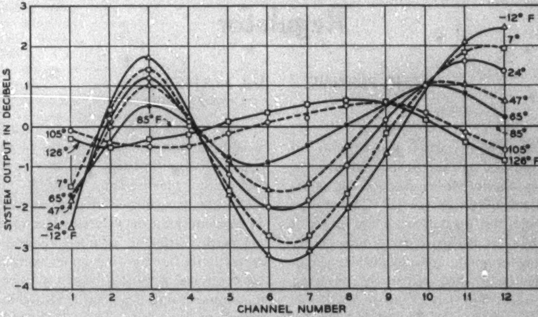


Fig. 3. Repeater gain change with operating temperature (12 repeaters in tandem)

for Vancouver, 150 degrees for Chicago, 125 degrees for New York, 80 degrees for New Orleans, and 50 degrees for Los Angeles. It is interesting to note that the smallest figures are for the California coast line, which has a range of only about 40 degrees Fahrenheit. An average value of about 100 degrees Fahrenheit seems to be a representative figure for the United States as a whole and has been chosen for this study.

The transmission change with temperature for the 19-gauge toll cable is shown on Fig. 2 for equal mileage of the two line frequency groups. The flat component of transmission change has been neglected as it is corrected for by the flat gain regulation in each repeater and thus is not cumulative. The change in loss is plotted as a function of channel number, not frequency, to reflect the partial canceling of the slope and bulge components. It may be noted that a 45-degree temperature change results in a net residual slope change of about 0.025 decibel (db) per mile and a net residual bulge change of about 0.003 db per mile.

The transmission variation with temperature for the repeaters is shown in Fig. 3. It is based on laboratory tests on 12 repeaters and is plotted as a function of the system channel number. The data are presented as departures from the value at 85 degrees Fahrenheit which is considered a representative ambient condition.

To analyze the transmission changes in a long system, the effect of the representative temperature swing of 100 degrees Fahrenheit has been calculated for a 200-

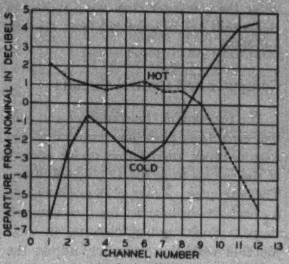


Fig. 4. 200-mile aerial cable system—deviation due to 50-degree change (hot or cold) from line-up temperature. Line-up temperature is 45 degrees Fahrenheit for cable and 85 degrees Fahrenheit for repeaters

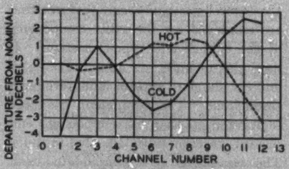


Fig. 5. 200-mile buried cable system—deviation due to 50-degree change (hot or cold) from line-up temperature. Line-up temperature is 45 degrees Fahrenheit for cable and 85 degree Fahrenheit for repeaters

mile system. These calculations have been made for both aerial and buried cable and the temperature change has been assumed to be 50 degrees both above and below line-up temperature. These variations are taken as typical of the magnitude of correction to be supplied by the regulator and are shown on Figs. 4 and 5. It will be noted that the requirements presented are based on information for channels 1 through 12. Since these requirements were formulated, channel 13 has been made available for general use in type-*N* systems as a substitute for a channel with excessive interference or marginal signaling. For long systems especially it became apparent that channel 13 was preferable to channel 1 from a transmission standpoint. Accordingly the regulator has been designed to use the channel carriers 2 to 13 inclusive. Considering Figs. 4 and 5, neglecting the transmission at channel 1, and projecting the curves to include channel 13, the maximum channel range requiring correction by the regulator is about 12 db at channel 13 for aerial cable and about 7 db at channel 13 for buried cable.

Description

The deviation regulator is a complete unit. It is connected into the transmission line directly following a low-high repeater. It provides shape changes only and has a negligible effect on the level of transmission. Each deviation regulator corrects for one direction of transmission so that two are required for both directions of a system. As each regulator employs 23 tubes it must be used at a power point.

The regulator consists of four main blocks which are shown schematically in Fig. 6. The regulator is shown connected into the transmission circuit between the low-high repeater output and the line. The four blocks of the regulator are indicated by the dotted lines which segregate the unit into the thermistor-controlled shaping networks, the regulator amplifier consisting of the line amplifier and pick-off amplifier, the channel filters, and the computer and control circuits.

Regulating Networks

The thermistor-controlled shaping networks are four separately controlled units. Each is continuously variable throughout its range under the control of its associated thermistor. The four network shapes are called slope, bulge, cubic, and quartic. The slope and bulge networks are basically required by the cable changes while the cubic and quartic networks compensate principally for the repeater changes. Flat regulation is provided by the low-high repeater and thus the regulator is relieved of this duty.

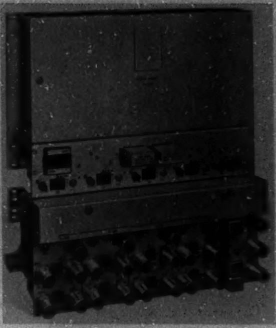


Fig. 7. N deviation regulator

As the thermistor resistance is changed the networks diminish the magnitude of their change from one extreme, pass through a mean condition where no shaping is introduced, and continue to an extreme inverse shaping. The slope network provides ±10 db of slope change. The other three networks provide only ±5 db of change.

Each network component acts independently and simultaneously with the others. This independency of action of the components obviates the need for having a different time constant for each shape and thereby permits the same type of thermistor unit to be used throughout.

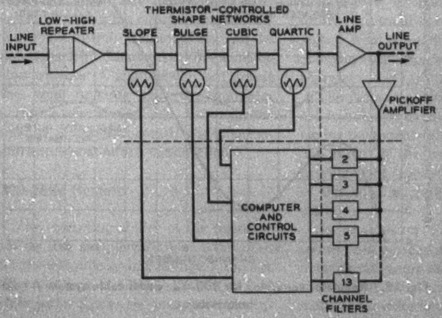
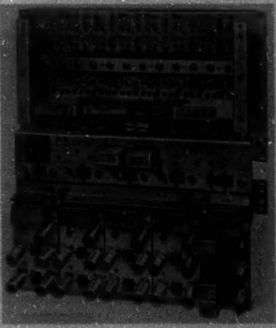


Fig. 6 (left). Regulator block schematic, showing by dotted lines the four basic circuit blocks of shaping networks, regulator amplifier, channel filters, and computer and control circuits

Fig. 8 (right). N deviation regulator with dust covers removed—the filters and computer are mounted at the top and plug-in units at the bottom



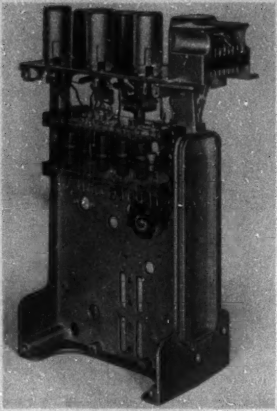


Fig. 9. Regulator amplifier—the mounting slide and terminal connector are visible at the right

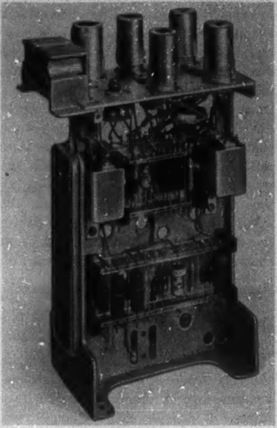


Fig. 10. Regulator control unit

Regulator Amplifier

The second box in the block schematic is the regulator amplifier which uses three *408A* vacuum tubes and is a plug-in device. It performs two functions and may be considered as composed of a line amplifier and a pick-off amplifier.

The line amplifier portion provides about 60-db gain to compensate for the mean loss of the regulating networks and thus restore the system signals to their proper level for transmission over the line. It is a 2-tube feedback amplifier very similar to the amplifier portion of the low-high repeater. It functions at the same signal levels and has similar transmission characteristics.

The pick-off amplifier portion provides a source of power to activate the control circuits. It is a 1-tube amplifier with local feedback to stabilize its gain. Its input is derived from a tap on the output transformer of the line amplifier.

Channel Filters

The output of the pick-off amplifier connects to the bank of 12 channel filters. These are the standard type-*N* channel band filters for channel carriers 2 through 13. At the output of each filter there is a potentiometer used primarily to adjust for manufacturing variations in the individual filters so that as the signals enter the computer they are a replica of those that are transmitted from the line amplifier.

Computer and Control Circuits

The function of the computer and control circuits is to deliver current to the four network thermistors under control of the signals from the 12 channel filters. The computer consists of 48 resistors, one for each channel carrier for each network shape. Four control circuits, in conjunction with the computer, adjust the thermistor currents until only a small residual regulation error remains.

Equipment Features

The regulator has been designed using the same basic equipment features that are employed for the other components of the type-*N* and type-*O* systems. It uses plug-in units, die cast frameworks, and the standard type of miniature parts throughout. The equipment has been segregated so that all components employing vacuum tubes are plug-in units. The remainder is a consolidated unit ready for relay rack mounting except for a small power panel which is completely detached.

The complete regulator, less the power panel, is pictured in Fig. 7, and in Fig. 8 with the dust cover removed. The regulating networks, channel filters, computing network, and miscellaneous apparatus are arranged on a filter and network panel. This panel is assembled with the type-*O* slide shelf at the bottom to mount the plug-in units. The regulator-amplifier unit is at the right side and the four identical control units are mounted to the left of it. These units use one of the basic type-*O* die-castings but with different face plates to accommodate the

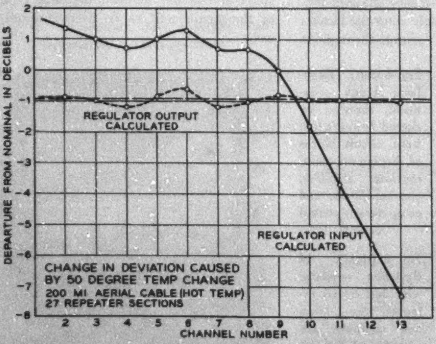


Fig. 11. Regulator correction for 200-mile aerial cable system at hot temperatures

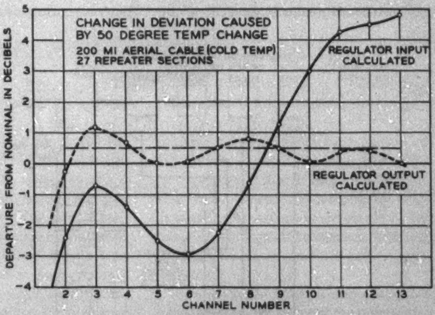


Fig. 12. Regulator correction for 200-mile aerial cable system at cold temperatures

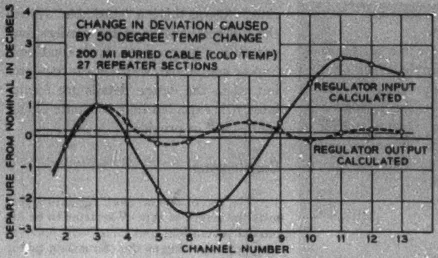


Fig. 13. Regulator correction for 200-mile buried cable system at cold temperatures

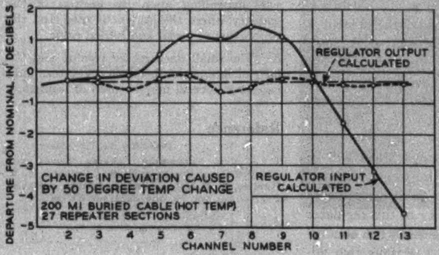


Fig. 14. Regulator correction for 200-mile buried cable system at hot temperatures

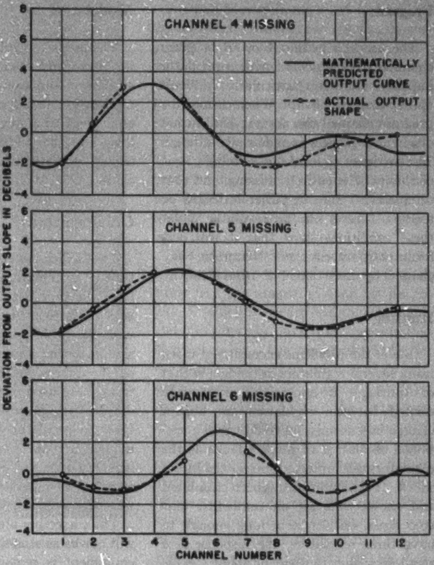


Fig. 15. Transmission change when one channel is removed

variations in tube and transformer requirements.

The regulator-amplifier unit is pictured in Fig. 9 and the control unit is shown in Fig. 10. These figures show the mode of construction and the accessibility of the parts for inspection or repair.

Performance

Extensive calculations and tests of the regulator performance have been made in the laboratory. The tests in every case verify the predicted performance within the accuracy of the experiment. Thus it appears that with any applied input characteristic within the capability of the regulator the output characteristic can be determined by calculation. The calculation consists of analytically performing the same function that the computer performs electrically and although the method is straightforward, it is laborious and requires accurate computation.

200-Mile System

The correcting ability of the regulator for the four sample variations established for the 200-mile system in Figs. 4 and 5 is shown in Figs. 11 through 14. In each of these latter figures the plot shows the input and output characteristic of the regulator. The input characteristic is that shown in Figs. 4 or 5. The output characteristic indicates that in each case good regulation has been achieved. The uncorrected residue in the output characteristic is small and for most of the channels is less than 1/2 db. The shape of this residue is beyond the ability of the four network shapes provided, and consists largely of higher order components. This conclusion is supported by the observation that the residue has five or six axis crossings. It will be noted that the average of the output characteristic does not fall precisely on the zero-db axis. This indicates a small flat gain component resulting from regulating action. This flat component is absorbed by the following repeater.

The excess regulating range in each of the four networks could be used to equalize for the manufacturing variations that will occur in the repeaters. The engineering rules required to take advantage of this equalizing ability are still under study.

An analysis of Figs. 11 through 14 shows that most of the slope and bulge network ranges are used, but only about a third of the cubic and quartic. The remainder provides an allowance for equalizing manufacturing variations.

The stiffness of this regulator is measured by the ratio of the amount of a given network shape in the input characteristic to the amount in the output characteristic. For these circuits this ratio is about 15. In the test measurements of the output characteristics this residual of the four corrective shapes is negligibly small compared to the magnitude of the higher order shapes.

Missing Channel

A circuit condition that is of particular interest in an operating system is the functioning of the regulator when a channel carrier is removed. This will happen whenever a channel unit is removed for test or repair. When such a condition prevails the regulator tries to correct for a very low carrier. The amount of change introduced into the transmission characteristic varies with the particular channel missing, but is a maximum of only 2 db in the worst case. Fig. 15 shows the calculated and observed performance for channels 4, 5, and 6, which is representative of all the channels. As this change is slower than the action of the individual channel regulator no serious system problem usually exists.

Program

The use of schedule-A or -B program channels over a system with a regulator imposes another particularly difficult operational problem. When using the program channel the normal distribution of channel powers is materially altered. The carrier of channel 6 is 5 db high, that of channel 7 is entirely missing, and that of channel 5 may be either 5 db low or missing. Tests on the regulator under these conditions show that it will perform satisfactorily, not distorting any channel by more than 1 db from ideal.

Hunting

One of the problems encountered when using several regulators in tandem is that of hunting. If appreciable interaction existed between regulators, an abrupt change in transmission could cause a sustained oscillation in shape change as the regulators attempted to correct the initial shock. This condition has been calculated and shows that for an impressed unit step signal on a system long enough to have three regulators, the response is a small decaying transient of about 1 cycle over a 20-minute period.

Modulation

The use of a regulator in a system does not have any measurable effect on the system modulation products. From a modulation standpoint the regulator is equivalent to a single repeater and for a long line the addition of one repeater should not increase modulation appreciably.

ON Regulator

The facilities offered by the type-N and type-O systems are being combined to form the type-ON system. This system uses 10 carriers instead of 12 but otherwise the cable portion of the system is similar to that for an N system. As some of the carrier frequencies are different from those used on type N, a somewhat different regulator is required. This new type-ON regulator will work in the same manner as that for type N but its regulating networks and channel filters will be adjusted to different frequencies, and it will have a slightly different computer.

The correcting ability of this regulator will not be as sensitive and when channel equipment is removed or fails two adjacent carriers may be lost at the same time. Also as this is a "stackable" system the full complement of ten carriers may not always be present. Such conditions impose special hardships on the regulator but when the design details are formulated it is expected that adequate circuit control will be provided.

Conclusions

1. The use of one or more regulators permits the length of type-N systems to be extended severalfold and effectively removes temperature swing as the controlling factor in limiting system length.

2. The line variations indicate that a regulator will be used at 100- to 200-mile intervals depending upon the section of the country where the system is used and the proportion of aerial and buried cable.

3. The small uncorrected residue and the lack of interaction between regulators indicate that several may be used in tandem.

Reference

1. A NEW TELEPHONE CARRIER SYSTEM FOR MEDIUM-HAUL CIRCUITS, R. S. Caruthers, H. R. Huntley, W. E. Kahl, Ludwig Pedersen. *Electrical Engineering*, vol. 70, Aug. 1951, pp. 692–97.

No Discussion

Heat Dissipation from Toll Transmission Equipment

J. A. COY
MEMBER AIEE

THE liberal use of vacuum tubes in toll transmission equipment, the miniaturization of this and other component apparatus, and the assembly of the components into cubic units rather than on flat panels, for the purpose of realizing the potential space saving, has presented a new problem in heat dissipation. The necessity of designing for safe heat disposal has long been recognized in radio transmitters and similar high-power equipment. Now it is apparent in low-power equipment as well.

The situation is typified by N carrier terminals,[1] but is not by any means limited to this type of equipment. The latest telegraph and telephone transmission equipments all present a similar problem, and this trend will continue and expand as miniaturization is extended and components become even smaller. Conversion from vacuum tubes to transistors will reduce the heat developed per individual circuit, but the transistors must be kept relatively cool, (140 degrees Fahrenheit (F) is a presently accepted limit except for phototransistors, for which the accepted maximum is 130 F). The rapid increase in number of circuits to meet the present demands for toll service, and to take care of dialing to replace manual operation, will tend to increase the total heat per office to be dissipated.

A rather vivid illustration is found in an installation of 120 N carrier terminals which can easily be contained in a room 20 by 30 feet. In such an installation 42 kilowatts of energy will be released, equal to that obtained from the combustion of 720 gallons of fuel oil per month, enough to heat two ordinary houses in midwinter. Several installations are now approaching this size, and space is being reserved, for as many as 300 terminals in one room.

This example and the methods of treating it will be discussed in some detail, since it is typical of the conditions which we can expect in the future. In Fig. 1 the heat generation in N carrier terminals and various other types of telephone equipment is compared on the basis of watts per square foot of assigned space, and also watts per cubic foot of room volume. The assigned space includes reasonably liberal allowance for good working aisle widths and room for storage lockers, work bench or desk, and similar accessories. It will be noted that the more recent developments, such as type-N and type-O carrier, $E2$ repeaters, the telegraph service board, and 2,600-cycle

---

Paper 53-344, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Pacific General Meeting, Vancouver, B. C., Canada, September 1–4, 1953. Manuscript submitted May 25, 1953; made available for printing July 8, 1953.

J. A. COY is with the Bell Telephone Laboratories, Inc., New York, N. Y.

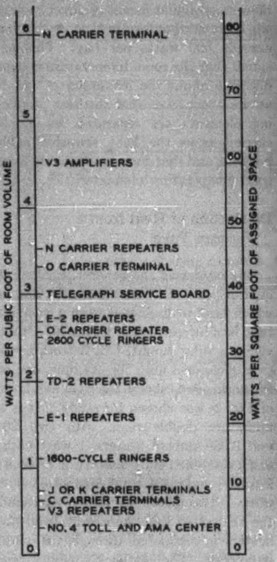


Fig. 1. Heat generation in toll transmission equipment

ringers are all heavy producers of heat, while the older types of toll transmission equipment, together with switching and associated equipment, are relatively light producers in terms of heat per unit space. Considered alone, *V3* voice-frequency amplifiers are heavy producers but they are likely to be associated with inert equipment to compose *V3* telephone repeaters, which appear further down the list.

The satisfactory disposal of the heat generated falls naturally into two stages. First, the heat must be removed from the point where it is generated, without impairing the operation or reducing the life of any component apparatus mounted in the same unit and, second, the total heat generated must be disposed outside the office without too much increase in the room temperature.

Removal of Heat from Hot Spots

In the *N* carrier terminal, nearly all of the heat originates in the vacuum tubes, and so these have been mounted on the front of the unit, as in Fig. 2, where the maximum amount of heat is carried away by conduction and radiation, and the minimum amount reaches the body of the unit. Within this body are mounted elements which do not produce heat, as in Fig. 3, including the thermistors and varistors whose operation is most affected by temperature. The units are plugged into a mounting framework to form a complete terminal as in Fig. 4, which shows the lowest of three terminals mounted in one relay rack bay.

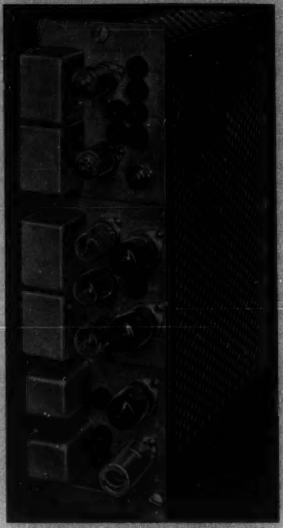


Fig. 2. Channel unit, front view

The design objective for the terminal was to keep the temperature of the critical elements such as the varistors of the compandors and modulators below a maximum of 140 F. This also was the temperature above which the potting compound then used for some of the transformers began to soften. Subsequently this was replaced with a compound good to 155 F. To facilitate heat dissipation, the tube shields were omitted from all but a few of the vacuum tubes, and the covers on the rear of the units were made of perforated metal. In the cover visible in Fig. 2, for example, 85 per cent of the sheet metal is removed. The temperature rise of critical apparatus above ambient was thus reduced by perhaps 10 F but it was still around 40 to 50 F, measured in an installation of only a few terminals. To insure adequate margin for large installations with lines of *N* carrier terminals back-to-back and facing one another across the aisle, and for operation in hot climates, it was decided to introduce bay blowers.

A small blower is mounted just below the bottom terminal in each bay as shown in Fig. 4. Air from this blower is conducted through two rectangular ducts up the sides of the bay and discharged in jets just below the concentrations of vacuum tubes which may be identified in groups of five near the center of each unit. These jet discharge openings may be seen in the left-hand duct. The velocity of the jets (2,700 to 4,100 feet per minute) is sufficient to spread the air well across the bay.

The blower delivers about 100 cubic feet per minute total for the bay, 5 to 8 cubic feet per minute in each of the 15 jets, of which there are 9 in the left-hand and 6 in the right-hand duct opposite the channel units. The temperature rise in the units is reduced by this means to half the amount which would occur with only natural cooling. With this arrangement, good operation of the equipment may be expected with an ambient temperature up to 105 F measured 6 feet above the floor in the center of the aisle in front of the *N* carrier terminals. Permanent injury of the equipment is unlikely if the room temperature should rise as high as 130 F with the blowers operating, or 100 F with the blowers stopped. With the adoption of the bay blowers, it was possible to revert to the use of solid, in place of perforated, rear covers for better mechanical and dust protection.

In order that the blower may operate only when needed, it is controlled by a thermostat located in the mounting shelf nearest the top of the bay. Fig. 5 shows the operating temperature of the apparatus inside the channel unit with the blower off, with the blower being controlled by the thermostat, and with the

Fig. 3. Channel unit, side view

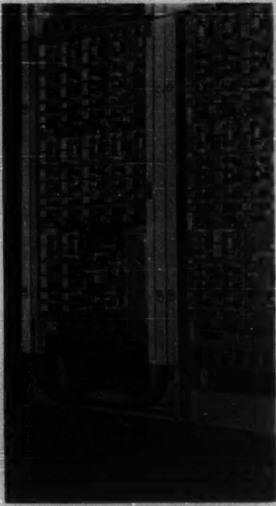


Fig. 4. N1 terminal, lowest unit of 3, with blower for cooling the bay

blower on continuously, plotted against the ambient temperature at the 6-foot level in the room. A temperature band is shown rather than a single line, to indicate the temperature spread within the units. The temperature measurements were made by placing thermocouples on the surface of several important elements. Since no heat was being generated in these elements the surface temperature so measured was the same as the interior temperature of the component.

If both the terminals and the blowers were operated only from commercial power supply, a power interruption would both stop the blowers and stop the generation of heat. In the usual case, however, the terminals are operated from storage battery reserve when power fails so that heat continues to be produced. In most of the larger offices, engine alternators are available for emergency supply, and the period of battery operation is short. In these offices transfer of the blowers to the emergency alternator is desirable. It does not appear necessary to provide such an emergency source for the blowers alone or to introduce the added complication of d-c operation from the batteries, since the rate of temperature rise in the room and in the equipment will be slow enough to cover a considerable period of power failure. Fortunately, long power outages are least likely in summer.

The *N* carrier repeaters generate only about half as much heat in a given space as the terminals, about 500 watts per relay rack bay for repeaters instead of 1,000 watts per bay for the terminals. When these repeaters are installed in buildings, blowers are not required because the temperature rise is only about half as much, and because the apparatus will operate well at somewhat higher temperature.

In general, it may be assumed that bay blowers or similar means of direct cooling can be avoided unless the heat production exceeds 500 watts per bay. This assumes that the room temperature is kept down to about the maximum of 105 F at the 6-foot level, that the heat producing elements are separated, as in *N* carrier, from the heat sensitive components, and that these components have good temperature characteristics.

Dissipation of Heat from a Vacuum Tube

In considering the general problem of heat disposition in toll offices, the way in which heat is dissipated from a vacuum tube is of interest. For this a *407A*-type vacuum tube mounted in a working *N* carrier channel unit, for example, one of the unshielded tubes in the right hand row of Fig. 2, was chosen for study. It is a miniature double-triode with two 20-volt 0.050-ampere heaters, 1 watt each, and a combined plate dissipation of about 2 watts, making a total dissipation of 4 watts. The plate structures surround the heaters so that the energy from the heaters is dissipated through the plate structures. The ambient temperature was 80 F, the glass envelope temperature measured with a thermocouple was 160 F. The plate temperature was estimated to

Fig. 6. Comparison of heat generated and heat dissipated from room

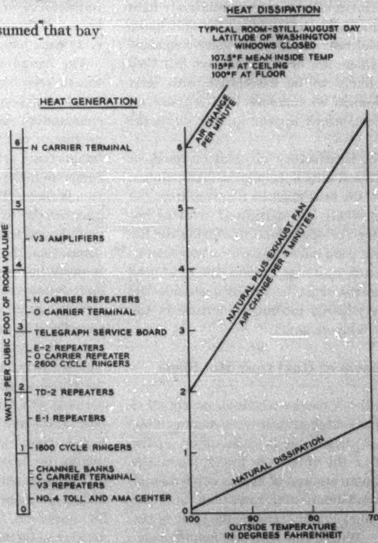
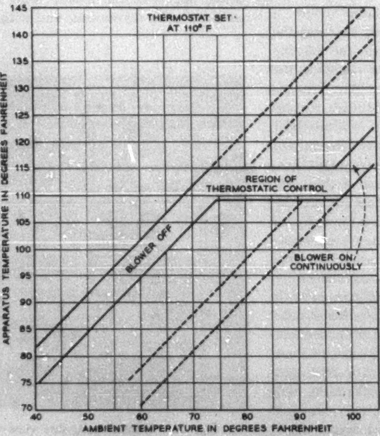
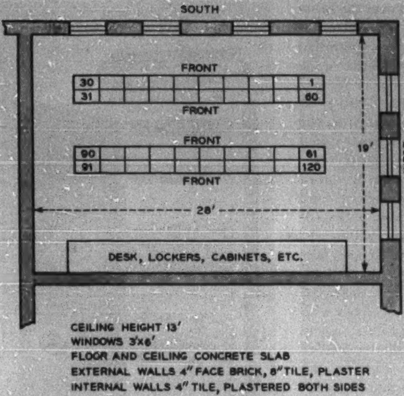


Fig. 5. Apparatus temperature in N1 carrier terminal




SOUTH
FRONT
30
1
31
60
FRONT
FRONT
90
61
91
120
FRONT
19'
WEST
28'
DESK, LOCKERS, CABINETS, ETC.
CEILING HEIGHT 13'
WINDOWS 3'x6'
FLOOR AND CEILING CONCRETE SLAB
EXTERNAL WALLS 4" FACE BRICK, 8" TILE, PLASTER
INTERNAL WALLS 4" TILE, PLASTERED BOTH SIDES

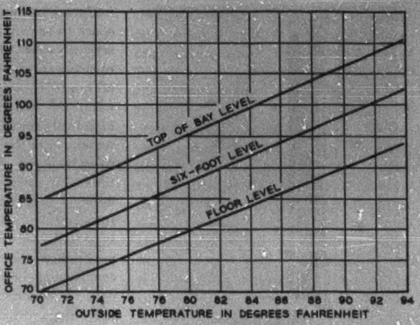


Fig. 7 (left). Floor plan, typical installation

Fig. 8. Observed temperatures in an office with approximately 50 changes of air per hour

be 480 F. From this the following data were calculated or estimated: Radiation from plates and envelope, 0.7 watt; convection from envelope, 0.3 watt; conduction through socket and wires, 0.3 watt. The conduction through the socket and wires was found to be surprisingly small. The temperature difference between them and the face of the unit or the connected apparatus was scarcely measurable.

Somewhat more than half of the radiation sphere around the tube is subtended by the panel and the transformers on the front of a channel unit like that shown in Figs. 2 and 3. Perhaps a quarter of the convected heat is transmitted to the face of the unit. All of the conducted heat enters the body of the unit. Of the total of 4 watts, then, from a *407A* tube, about 2¹/₄ watts reach the body of the channel unit. If it is assumed that the heat from a *408A* tube (2 watts) is distributed in the same way as that from a *407A* vacuum tube, then of the 24 watts generated in all the tubes of a channel unit, about 13.5 watts will reach the body of the unit.

The heat assumed to have thus entered the body of the channel unit must be dissipated from the front and rear areas since, as is apparent in Fig. 4, each unit is surrounded on the top, bottom, and sides by other units. From the front of the unit with an effective area of 70 square inches and a measured temperature rise of 45 F, the dissipation by radiation and convection is 11 watts. From the rear with equal area and a 20-degree rise, the dissipation is 3 watts. A small dissipation occurs through the framework, probably less than 1 watt. This total of about 15 watts is in fair agreement with the 13.5 watts estimated in the foregoing to be transmitted from the tubes to the unit.

Removal of Heat from Offices

The heat generated in the equipment must eventually be removed outside the building. There are three possible ways of doing this: first, by natural dissipation; second, by forced circulation of outside air through the room by means of an exhaust fan or a combination of exhaust and intake fans and, third; by refrigerative cooling. The method and extent of action to be taken will depend on the amount of heat generated per unit volume of room space; the type, arrangement, and usage of the building; and the climate; the final criterion, of course, being the relative cost.

The heat generation of *N* carrier terminals and various other types of equipment, as shown in Fig. 1, is repeated on the left-hand side of Fig. 6, for comparison with heat dissipation from the room, shown on the right-hand side of Fig. 6. The particular installation of *N* carrier terminals selected for this study is that of Fig. 7.

Obviously, natural heat dissipation becomes ineffective as the outside temperature approaches the maximum temperature to be allowed within the office, particularly when the windows are closed. This is indicated by the lowest curve on the right-hand side of Fig. 6. The curves in this figure are based on the room of Fig. 7, with two outside walls of substantial brick construction, insulated, and with windows as indicated. A still August day is assumed in the latitude of Washington, D.C., closed windows and room temperatures of 100 F at the floor, 115 F at the ceiling, and 107.5 F mean.

More effective heat removal can be provided by installing an exhaust fan to take the hot air out of the room and an intake to permit the entrance of the relatively cooler air from outside. Dust eliminating filters and a fan at the intake are advisable to avoid aggravating housekeeping problems. Two curves on the right-hand side of Fig. 6 indicate the results to be expected from forced air circulation the lower where the air flow is such that the air in the room is completely replaced every 3 minutes, the other where it is replaced once a minute.

Recently, an office with seventy operating terminals was equipped with intake fans to draw in the air through filters, and distribute it by means of ducts at floor level along the sides of the room, and an exhaust fan to draw air from the ceiling level and discharge it outside the building through a plenum chamber treated to reduce noise. Thermostatic control regulates the proportion of outside air circulated to room air recirculated. Temperatures were measured from June to September in this office at floor level, at a level of 6 feet above the floor, and at the top-of-bay level, approximately 12 feet above the floor, with an estimated 50 changes of air per hour. From these data, the curves of Fig. 8 were plotted. In this room with approximately 7,000 cubic foot volume, 70 terminals were in operation, dissipating 24.5 kw or 3.5 watts per cubic foot. The observed data check the theory about as well as could be expected in view of the difficulty of distributing the air efficiently at this high rate and of obtaining temperature measurements that

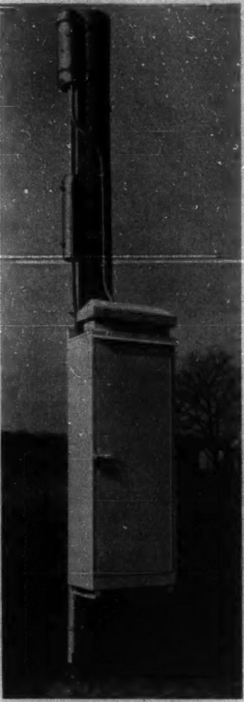


Fig. 9. Pole-mounted N1 repeater cabinet

represent average conditions. Probably the most interesting result is the lack of any discomfort with such a high rate of air circulation. The reason for this, no doubt, is that the circulated air is so near body-surface temperature. Careful engineering is needed in the introduction and distribution of air at these high rates so as to obtain as uniform conditions as are practicable throughout the room.

As compared with a ventilating system having a high rate of air change with consequent heavy dirt infiltration, the economies of refrigerative air conditioning may well be considered. The consideration of such a system will include temperature and humidity control. The air is filtered, cooled, and dehumidified to the required degree in a common equipment, from which it is circulated through the areas to be conditioned and returned for recirculation. A proportion of outdoor air is introduced for freshening. Installations of this type are generally engineered to provide a fairly uniform distribution of air over the entire space to be conditioned. The balance is seriously upset when a large concentration of heat is introduced in a small part of the area. It is then generally preferable, where this can be done, to partition off high-heat-producing equipment from the remainder of the office or to locate it in a room by itself. Where general air conditioning is available this can be drawn upon for part of the cooling required and supplemented as needed by packaged-type unit air conditioners within the partitioned space. Where general air-conditioning is not available, packaged-type air conditioners can be used for the full duty and located in the room with the heat-producing equipment. Where the primary objective is the removal of heat generated by the equipment, the unit conditioners can be operated with only thermostatic control, since in this case humidity control is not necessary.

If it is assumed as a first approximation that just enough refrigerating capacity is provided to care for the heat generated in the equipment, then the refrigerative capacity for 120 terminals, for example, located where the heat conditions are as represented in Fig. 6, will be approximately 12 tons, 1 ton of refrigerating capacity being defined as the heat required to freeze 1 ton of water at 32 F per 24 hours, or 12,000 British thermal unit per hour. The cost of such refrigerative equipment may amount to a little more than the cost of the building space to house the equipment or, in the case of *N* carrier terminals, to perhaps 2 per cent of the cost of the equipment itself. It may actually be cheaper than high-rate outside air circulation where this requires extensive duct work.

New Equipment Arrangements

It has frequently been stated as an axion that heat generated in equipment should be removed at its source and not released into the room to be taken care of by space air conditioning. Examination of specific methods, however, does not prove that this is necessarily axiomatic. Too often in considering relative methods the objective can be inadvertently changed so that one method is considered only so far as equipment safety is affected, while another is made to include employee comfort.

Various methods have been suggested, such as, for example, turning a pair of back-to-back equipment bays inside out so that the vacuum tubes, instead of being in the aisles, project into a duct space between the two bays. A refrigerant could then be run through a finned pipe in the base of the bays and cold air circulated up the chimneylike central space by a small exhaust blower at the top. Other possible arrangements involve the circulation of a refrigerant directly through the equipment to be cooled.

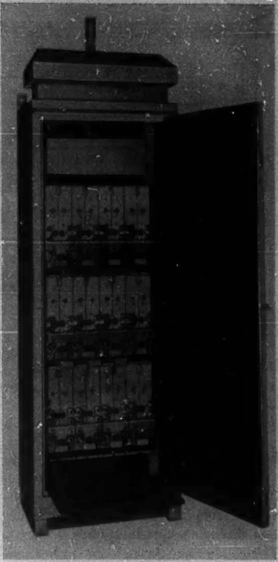


Fig. 10. Arrangement of 12 N1 repeaters in a pole-mounted cabinet

While such arrangements are obviously practicable, they generally involve relatively expensive means for the circulation of the refrigerant and the air, the presence of refrigerants may increase the hazard to service, and the component apparatus may be made less accessible for operation and maintenance. Fitting them into existing offices, existing methods of manufacture, and (perhaps even more) into the engineering consciousness of those who are responsible for equipment design and office layouts, would constitute a major undertaking which does not seem to be justified by the advantages, except in unusual situations where the heat generated cannot readily be taken care of otherwise.

Pole-Mounted Repeater Cabinets

At many of the repeater points in *N* carrier systems the repeaters are located in cabinets mounted in the open on line poles as in Fig. 9, or on a cross arm between the line pole and a stub. Each cabinet contains 12 2-way repeaters arranged as shown in Fig. 10. Power is fed to these repeaters from the next adjacent office over the same cable pairs used for transmission. No commercial power supply is needed therefore at these pole-cabinet locations. As a result, no power is readily available for the operation of ventilating blowers. A full complement of 12 repeaters dissipates approximately 300 watts. The requirement set up when the development was initiated, was for operation over a climatic temperature range of −40 F to +120 F.

The cabinet is lined with 1-inch-thick Cellotex for insulation against winter cold. A ventilator is provided at the top and an opening in the bottom for summer cooling by natural convection. The top opening is controlled by a thermostatically operated shutter. With this shutter closed the heat dissipation from the insulated cabinet in still air is calculated to be 1½ watts per degree of temperature difference between the interior and exterior. If the cabinet is only partially equipped, say with 5 repeaters, the 125 watts of energy dissipated would, under ideal conditions, raise the temperature 80 F. Even with liberal allowance for air leakage, effect of wind, and other variable conditions, the rise should be at least half the theoretical value, keeping the inside temperature above zero degrees F. The repeaters will operate satisfactorily at this temperature or even lower. More repeaters in the cabinet will of course, raise the temperature higher. An initial installation of fewer than five repeaters is unlikely for any considerable period.

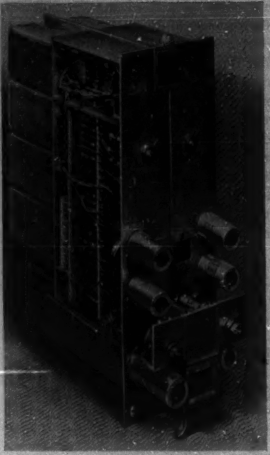


Fig. 11. N1 carrier repeater

As the temperature in the cabinet is raised the thermostat will begin to open the shutter in the top ventilator at about 60 F and will insure full opening at some temperature below 100 F. As shown in Fig. 11, the repeaters are arranged as are the terminal units, in that the vacuum tubes and power adjusting resistors are placed on the face of the units, and apparatus components not producing heat, inside. They are placed in the cabinet so that the heat-producing elements are located in a chimneylike compartment in the front of the box, Fig. 12. The main air flow is up this front space, while a secondary flow takes place between the rear of the repeater units and the back of the box. Measurements made in the laboratory indicated that the temperature rise of apparatus inside the repeater units in the top row would be about 35 F. Measurements made on working installations outdoors under various summer conditions gave results more variable than had been expected, as indicated in Fig. 13. The difference probably is largely because of the effect of wind, particularly of variable light breezes, in damping out the natural chimney effect. The apparatus temperature rise, as might be expected, varies almost directly with the temperature of the air issuing from the ventilator. Fig. 13 indicates that the usual temperature rise to be expected will be about 50 F. The apparatus will sustain no permanent damage below about 160 F, and up to this limit the operation of the repeaters will not be affected. This would indicate that safe operation may be expected in ambient temperatures up to 110 F and, because of the time lag between outside temperature rise and apparatus temperature rise, this limit could be exceeded for short periods. While the ideal of +120 F atmospheric temperature operation has not been met, it has been approached so closely that only a few very bad locations will need special attention. Weather bureau records indicate that temperatures above 110 F are seldom reached in

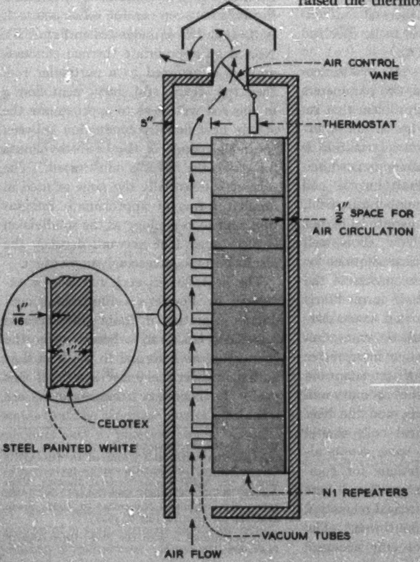
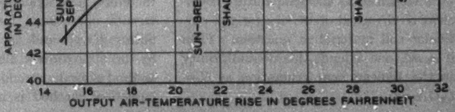


Fig. 12 (left). Temperature control in pole-mounted cabinet

Fig. 13 (below). Observed apparatus temperature rise in a pole-mounted cabinet



this country except in the southwestern deserts.

Conclusion

The disposition of heat from present day miniaturized equipment requires attention from the design engineer. Hot-spot temperatures in the units need to be kept below the limit where components will be damaged or operation impaired. Basically this requires simply that the heat be removed rapidly enough from the immediate vicinity where it is generated. The problem of accomplishing this, however, usually is resolved into two parts, first removing the heat from each unit and, second, disposing of it outside the office. The second problem is of a type readily soluble by ordinary heating and air-conditioning methods. Apparatus arrangement, air circulation, and, where necessary, refrigerative cooling are the means available. The introduction of transistors in place of vacuum tubes will require new methods of solving the first phase and will reduce the total heat to be disposed of in the second. Continued growth of demand for facilities, however, will enlarge the over-all problem.

Reference

1. A NEW TELEPHONE CARRIER SYSTEM FOR MEDIUM-HAUL CIRCUITS, **R. S. Caruthers, H. R. Huntley, W. E. Kahl, Ludwig Pedersen.** *Electrical Engineering*, vol. 70, Aug. 1951, pp. 692–97.

No Discussion

Temperature Fields in Electrical Coils: Numerical Solutions

P. J. SCHNEIDER
NONMEMBER AIEE

Synopsis: This paper illustrates the computation of steady temperature fields in electrical coils by the numerical-relaxation method of Emmons and Southwell. The general numerical method is briefly reviewed, and several examples serve to exemplify the simplicity of the technique and to verify that numerical temperatures are in substantial agreement with both experimental data and exact analytical values.

A PLANE cutting normal to the curved longitudinal axis of a typical electrical coil of toroidal form reveals a profile (cross section) of insulated metallic wires. For analytical purposes one replaces this actual coil by a mathematical model whose longitudinal axis has a large radius of curvature, and whose profile is a homogeneous area of uniform thermal conductivity k. Higgins[1] gives a comprehensive outline of the physical basis for this mathematical model. If the coil carries constant electric current i and is of electrical resistivity ρ, then the Joulean heat generated within the coil is $Q=i^2\rho$ at a profile temperature t, and the heat developed at a reference temperature $t=0$ is correspondingly $Q_0=i^2\rho_0$. In the temperature range of normal coil operation, ρ varies linearly with local profile temperature as $\rho_t=\rho_0(1+\alpha_0 t)$ wherein α_0 is identified as the temperature coefficient of electrical resistivity for the particular coil material in question. Then if $\alpha_0\neq 0$, one extends this nonuniformity to the Joulean heat generation as $Q_t=Q_0(1+\alpha_0 t)$.

The mathematical model of the electrical coil effectively focuses attention on the steady-state 2-dimensional temperature distribution in a region generating nonuniform heat and having the same boundaries as the profile of the prototype coil. All heat conducted to the profile boundary is assumed to be dissipated to an ambient medium of uniform temperature t_a, where the unit thermal conductance on this boundary is to be specified by a uniform coefficient h_s.

Having identified the shape and dimensions of the profile and the parameters Q_0, α_0, k, t_a, and h_s, the problem that remains is to find $t(x,y)$ by the most expeditious method. Such information is needed to predict the maximum permissible electrical loading limit for the coil as based on this general temperature field, and the location and magnitude of hot spots in the profile. Now, it is well known that exact analytical solutions for $t(x,y)$ are available if the profile of the coil is of simple geometric form. It is also common experience that exact solutions for the more realistic cases frequently require the excessive interpretive and computational labor attendant to infinite trigonometric series, ordinary and modified Bessel functions, and the like. The designer of electrical coils should therefore have available some simple alternate operational technique for computing $t(x,y)$, and such is offered here by the approximate numerical-relaxation method of Emmons and Southwell.[3] The numerical method is not only accurate enough to be competitive with exact solutions (if available), but flexible enough to encompass a number of practical cases for which exact solutions are intractable. One of several possible experimental approaches to this problem is contained in the literature.[3,4]

The Numerical Method

For purposes of illustration, consider a general region of uniform thermal conductivity and developing distributed linear heat as in Fig. 1(A). The first step in a numerical solution for $t(x,y)$ is to subdivide this system into a number of small regular subvolumes. The subdivision (square in this case) is then replaced by a network of fictitious heat-conducting rods between central nodal points 1, 2, 3, 4, and 5, as indicated in Fig. 1(B). Now if an appropriate thermal conductance K is assigned to a particular rod, then the steady and linear heat flow q in this rod will serve to approximate the steady heat flow by conduction between the nodal points of the two subvolumes for which the rod was substituted. The scheme is essentially the same as used in familiar forms of approximate integration, and so in analogy as the subdivision is made finer (the network denser), the results become increasingly more exact.

The heat flux in each rod is now expressed by Fourier's 1-dimensional conduction equation in finite differences as $\Delta q=K\Delta t=k(a/L)\Delta t$, wherein a is the conducting area normal to the heat flow path and L is the linear length of this path. If boundary temperatures are

---

Paper 53-352, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Pacific General Meeting, Vancouver, B. C., Canada, September 1–4, 1953. Manuscript submitted November 10, 1952; made available for printing July 2, 1953.

P. J. SCHNEIDER is with the State University of Iowa, Iowa City, Iowa.

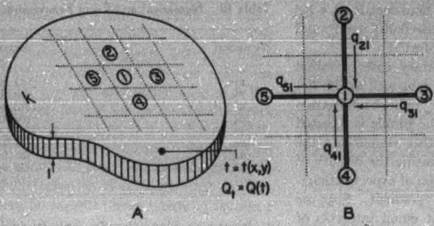


Fig. 1 (left). Subdivision and network of heat-conducting rods for numerical solution

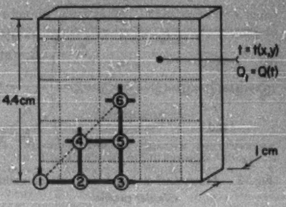


Fig. 2 (right). Numerical network for coil of solid square profile

unknown, then one also places conducting rods in the boundary and this heat loss is expressed by Newton's equation as $\Delta q = h_s A \Delta t$ in which A is the boundary area associated with each boundary nodal point. Now, the steady heat flow by conduction into a typical nodal point 1 of volume v_1 feet³ from all four adjacent nodal points 2, 3, 4, and 5, together with the heat generation at 1 in amount Q_t Btu per hour-feet³, is given by $\bar{q}_1 = Q_{t_1}v_1 + q_{21} + q_{31} + q_{41} + q_{51}$ as

$$\bar{q}_1 = Q_{t_1}v_1 + K_{21}\Delta t_{21} + K_{31}\Delta t_{31} + K_{41}\Delta t_{41} + K_{51}\Delta t_{51}$$

or

$$\bar{q}_1' = \frac{Q_{0_1}v_1}{k} + \frac{Q_{0_1}v_1\alpha_0}{k}t_1 + (a/L)_{21}(t_2 - t_1) + (a/L)_{31}(t_3 - t_1) + (a/L)_{41}(t_4 - t_1) + (a/L)_{51}(t_5 - t_1)$$

where for the special case of uniform square subdivision the a/L's $= 1$.

Then, in general, for a subdivided region of n nodal points, one writes the set of n simultaneous algebraic residual equations

$$\left.\begin{aligned} \bar{q}_1' &= \frac{Q_{0_1}v_1}{k} + \frac{Q_{0_1}v_1\alpha_0}{k}t_1 + \sum_{\alpha}(a/L)_{\alpha 1}(t_\alpha - t_1) \\ \bar{q}_2' &= \frac{Q_{0_2}v_2}{k} + \frac{Q_{0_2}v_2\alpha_0}{k}t_2 + \sum_{\beta}(a/L)_{\beta 2}(t_\beta - t_2) \\ &\;\;\vdots \\ \bar{q}_n' &= \frac{Q_{0_n}v_n}{k} + \frac{Q_{0_n}v_n\alpha_0}{k}t_n + \sum_{\lambda}(a/L)_{\lambda n}(t_\lambda - t_n) \end{aligned}\right\} \tag{1}$$

In the steady state all residuals $\bar{q}_i'$ must be zero above the datum $Q_{0_i}v_i/k$, and this is in simple analogy to the Kirchhoff electrical circuit law which states that the algebraic sum of all steady currents at a junction must vanish. The remainder of the solution is concerned with finding all t_i that satisfy equations 1, and this can be accomplished by any one of the available methods of solving a system of n simultaneous algebraic equations in n unknowns. However, when n is large, such methods as solution by determinants become somewhat tedious, and this is particularly so in cases where the residual equations have nonintegral coefficients. Southwell[5] has developed an alternate relaxation method for this portion of the work, and rather than explain his method here we can go on to a number of examples in which the purely formal procedure of a relaxation solution will become evident.

An Example Checked by Experiment

Higgins[1] gives data from electrical-coil experiments by Rogowski and Vieweg[6] in which the isothermal boundary of a toroidal coil of square profile, 4.4 by 4.4 centimeters (cm) was maintained at 76.1 degrees centigrade (C), and the maximum profile temperature t_m recorded as 122.4 C. Higgins takes the Joulean heat generation in the coil as the linear function $Q_t = Q_0(1 + \alpha_0 t)$ and then computes the unit-volume heat development at $t = 0$ as $Q_0 = 0.1207$ watt per cm³, and selects the temperature coefficient as $\alpha_0 = 0.00423$ per C. Jacob[7] computes an equivalent uniform thermal conductivity for this particular coil as $k = 0.00531$ watt per cm, C. With these data one can easily effect a numerical solution and then compare the two t_m's.

Let the 4.4-cm² profile be subdivided as in Fig. 2 with half subvolumes at the boundaries. With the same uniform boundary temperature on all four edges, the temperature field has one-eighth symmetry so that only one-eighth of the profile need be considered as long as one remembers to double the effect of, say, rod 5–4 on the nodal point 4, and so on. Moreover, with isothermal boundaries the surface rods 1–2 and 2–3 are inactive. Then, for a typical internal nodal point such as 4

$$\bar{q}_4 = Q_{t_4}v_4 + 2K_{24}(t_2 - t_4) + 2K_{54}(t_5 - t_4)$$

But for all rods in a square network $K_{ij} = k(4.4/4)/(4.4/4) = k$, and for all 4.4/4 cm³ subvolumes of unit depth $Q_{0_i}v_i = 0.1207 \times (1.1)^2 = 0.146$, so that

$$Q_{t_i}v_i = 0.146(1 + 0.00423t_i), \text{ and}$$

$$\bar{q}_4' = 2t_2 + 2t_5 + 0.146/0.00531 - (4 - 0.146 \times 0.00423/0.00531)t_4$$

In like manner the three required residuals corresponding to equations 1 appear as

$$\left.\begin{aligned} \bar{q}_4' &= 2t_2 + 2t_5 + 27.5 - 3.884t_4 \\ \bar{q}_5' &= t_3 + 2t_4 + t_6 + 27.5 - 3.884t_5 \\ \bar{q}_6' & \;\; 4t_5 + 27.5 - 3.884t_6 \end{aligned}\right\} \tag{2}$$

The most elementary approach to finding the t_4, t_5, and t_6 that satisfy equations 2 is to venture a guess at these unknown temperatures and then compute the $\bar{q}_4'$, $\bar{q}_5'$, and $\bar{q}_6'$. In all probability these $\bar{q}'$'s will not be zero (above 27.5) as required, and consequently the t's would have to be successively changed until this condition was fulfilled. The Southwell relaxation procedure is a convenient tabular means of carrying out these successive guesses and cumulative corrections.

The relaxation of equations 2 was carried out for this example in Table I. The procedure was begun by first assuming all $\bar{q}_i'$ and t_i as zero on line 1. All $\bar{q}_i'$ were then corrected for the generated heat at each subvolume and according to what change would occur in each residual with changing the boundary temperature from zero to its correct value of 76.1 C. Thus in changing t_2 by 76.1, a change in $\bar{q}_4'$ in amount 2(76.1) = 152.2 is realized. Adding 27.5 to this value gives 179.7 as the corrected $\bar{q}_4'$, and this result together with similar corrections for $\bar{q}_5'$ and $\bar{q}_6'$ are entered on line 2 of Table I.

Table I. Relaxation of Residual Equations 2

	$\bar{q}_4'$	t_4	$\bar{q}_5'$	t_5	$\bar{q}_6'$	t_6
1	0.0	0.0	0.0	0.0	0.0	0.0
2	179.7	0.0	103.6	0.0	27.5	0.0
3	−388.4	100.0	200.0			
4	−208.7	100.0	303.6			
5	220.0		−427.0	110.0	440.0	
6	11.3		−123.4	110.0	467.5	
7			120.0		−466.0	120.0
8			−3.4		1.5	120.0
9	−13.6	3.5	7.0			
10	−2.3	103.5	3.6			
11	3.0		−5.8	1.5	−6.0	
12	0.7		−2.2	111.5	7.5	
13			2.0		−7.8	2.0
14			−0.2		−0.3	122.0

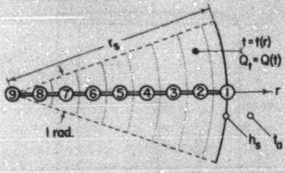


Fig. 3. Numerical network for coil of solid circular profile

One now advances the guess that, say, $t_4=100$, which effectively changes $\bar{q}_4'$ by $-3.884(100)=-388.4$. A change in t_4 in amount 100 also changes $\bar{q}_5'$ by $2(100)=200$. These changes are entered in Table I on line 3 and their algebraic sums are on line 4. The tabulation of guesses and corrections is carried on in this manner until all residuals in the network are reduced as nearly to zero as accuracy requires. In Table I the algebraic sum of the remaining residuals over the entire field is only 0.2, and to improve this further would require temperature changes of less than 0.5 C. If the end temperatures in Table I are correct, then the heat generated in the internal subregions should just balance the heat transferred to all four edges:

$$\Sigma q_0(1+\alpha_0 t)=\Sigma k\Delta t$$

$$\begin{aligned}&4\times 0.146(1+0.00423\times 103.5)+4\times\\&\quad 0.146(1+0.00423\times 111.5)+1\times\\&\quad 0.146(1+0.00423\times 122.0)\ 8\times\\&\quad 0.00531(103.5-76.1)+4\times\\&\qquad\qquad 0.00531(111.5-76.1)\end{aligned}$$

$$1.920\text{ watt}\doteq 1.917\text{ watt}$$

The maximum temperature in this coil computed by the numerical method is thus $t_m=122.0$ C to the nearest 1/2 degree. Despite the choice of a relatively crude numerical subdivision, this result is in close agreement with the experimentally measured value of $t_m=122.4$ C.

An Example Checked by Theory

To compare numerical results with an exact analytical solution for $t(x,y)$, consider the coil of solid circular profile dissipating heat to an ambient medium of uniform temperature $t_a=0$ degrees Fahrenheit (F). In particular, let $r_s=2$ inches, $k=1$ Btu per hour-feet-F, $Q_0=10^4$ Btu per hour-feet³, $\alpha_0=0.001$ per F, and $h_s=15$ Btu per hour-feet²-F.

This is an elementary case of radial conduction in which t_s as well as the $t(r)$ are unknown. The radial subdivision of any segment of the profile is selected with the view of either simplifying the computation or increasing the density of the network in a region of expected maximum temperature gradient. Suppose that temperatures at equal intervals of radius are required and, accordingly, choose the uniform radial subdivision in Fig. 3, where the bounding radii are separated by one radian in order to eliminate π from the computations.

Table II. Conductances and Volumes

$K_{1a}=2.5$	$v_1=0.3347/144$
$K_{21}=7.5$	$v_2=0.4375/144$
$K_{32}=6.5$	$v_3=0.3750/144$
$K_{43}=5.5$	$v_4=0.3125/144$
$K_{54}=4.5$	$v_5=0.2500/144$
$K_{65}=3.5$	$v_6=0.1875/144$
$K_{76}=2.5$	$v_7=0.1250/144$
$K_{87}=1.5$	$v_8=0.0625/144$
$K_{98}=0.5$	$v_9=0.0781/144$

The conductances and volumes appropriate to this subdivision are contained in Table II. As examples in their computation: $K_{1a}=h_s(2/12)=2.5$; $K_{32}=k(a/L)_{32}$ where $a_{32}=1.625/12$ and $L_{32}=0.25/12$, whence $K_{32}=6.5$; and the volume of region 3, $v_3=[(\pi/4)(3.25/12)^2-(\pi/4)(2.75/12)^2]1/2\pi=0.3750/144$. From Table II the nine required residual equations follow directly as

$$\left.\begin{aligned}\bar{q}_1'&=2.5t_a+7.5t_2+16.819-\ \ 9.983t_1\\\bar{q}_2'&=7.5t_1+6.5t_3+30.382-13.970t_2\\\bar{q}_3'&=6.5t_2+5.5t_4+26.042-11.974t_3\\\bar{q}_4'&=5.5t_3+4.5t_5+21.702-\ \ 9.978t_4\\\bar{q}_5'&=4.5t_4+3.5t_6+17.361-\ \ 7.983t_5\\\bar{q}_6'&=3.5t_5+2.5t_7+13.021-\ \ 5.987t_6\\\bar{q}_7'&=2.5t_6+1.5t_8+\ \ 8.681-\ \ 3.991t_7\\\bar{q}_8'&=1.5t_7+0.5t_9+\ \ 4.340-\ \ 1.996t_8\\\bar{q}_9'&=0.5t_8\qquad\qquad +\ \ 0.542-\ \ 0.499t_9\end{aligned}\right\}\quad(3)$$

Table III. Numerical and Exact Temperatures

Numerical	Exact
61	61.09
79	79.05
95	94.76
108	108.13
119	119.20
128	127.73
134	133.86
138	137.56
139	138.82

An exact solution for this case is given in the Appendix as

$$\begin{aligned}t&=\frac{(1/\alpha_0+ta)}{J_0(\sigma_s)-\mu J_1(\sigma_s)}J_0(\sigma)-\frac{1}{\alpha_0}\\&=1138.819J_0(\sigma)-1{,}000\qquad(4)\end{aligned}$$

wherein the J_n are first-kind Bessel functions of order n, $\mu^2=Q_0\alpha_0k/h_s^2$, and $\sigma^2=(Q_0\alpha_0/k)r^2$. Table III lists the approximate numerical temperatures to the nearest degree obtained through relaxation of equations 3, and the corresponding exact temperatures computed by equation 4. It is apparent that the two solutions are nearly identical.

Internally Cooled Coil

The hollow-profile internally cooled coil is a rather recent innovation for use in high-output generators. Consider the special case of a hollow-square profile (4 inches on a side and 1-inch thick walls) exposed to ambient air at 80 F and cooled

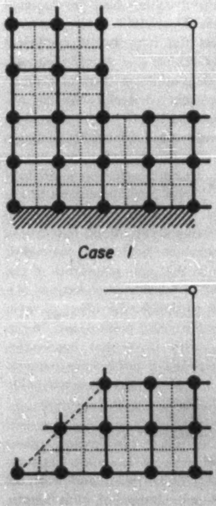


Fig. 4 (left). Numerical network for coil of hollow square profile

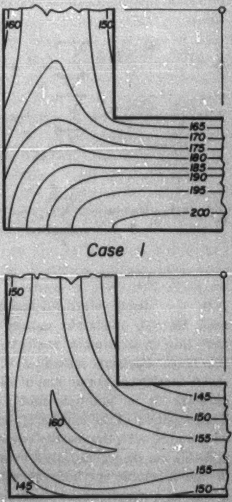


Fig. 5 (right). Temperature fields in internally cooled coils of hollow square profile

internally by a forced gas flow at a mean temperature of 120 F. Let $k=1$ Btu per hour-feet-F; $Q_0=86.4$ Btu per hour-feet^3; $\alpha_0=0.002$ per F, and select an outer unit surface conductance of 6 Btu per hour-feet^2-F, and an inner unit surface conductance of 24 Btu per hour-feet^2-F.

We consider two cases. In case 1 let a number of such coils be in contact with each other so that all contact surfaces can be regarded as adiabatic. In case 2 let the individual coils be separated so that all outer surfaces are exposed to ambient air. Temperature symmetry can be observed in both cases, and a uniform square subdivision selected as in Fig. 4.

The final relaxation of these systems gives the two distinctly different temperature fields shown in Fig. 5. These are solutions to an otherwise intractable problem by a method whose simplicity rests rather uniquely on a little algebra, and the elementary operations of addition and multiplication.

Appendix. Derivation of Equation 4

The symmetrical temperature field $t(r)$ for the circular profile of uniform thermal conductivity and generating Joulean heat as a linear function of local profile temperature is that which satisfies the ordinary differential equation

$$\frac{d^2t}{dr^2}+\frac{1}{r}\frac{dt}{dr}+\frac{Q_0}{k}(1+\alpha_0 t)=0 \tag{5}$$

and the particular boundary conditions for this case

$$\left(\frac{dt}{dr}\right)_s=-\frac{hs}{k}(t_s-t_a) \tag{6}$$

Defining a new dependent variable T as

$$T=\frac{Q_0}{k}(1+\alpha_0 t)=\zeta+\xi t$$

forces equation 5 to reappear as the familiar Bessel equation of zero order

$$r^2\frac{d^2T}{dr^2}+r\frac{dT}{dr}+\xi r^2T=0$$

whose general solution reads

$$T=C_1J_0(\sqrt{\xi}r)+C_2Y_0(\sqrt{\xi}r)$$

The second arbitrary integration constant C_2 is evaluated by noting that the Neumann second solution Y_0 cannot be retained for this case since T must remain finite as $r\rightarrow 0$. Thus with $\sigma=\sqrt{\xi}r$

$$t=\frac{C}{\xi}J_0(\sigma)-\frac{\zeta}{\xi} \tag{7}$$

The remaining integration constant is evaluated by the boundary condition of equation 6, according to

$$\left(\frac{dt}{dr}\right)_s=-\frac{C}{\sqrt{\xi}}J_1(\sigma_s)=-\frac{hs}{k}\left[\frac{C}{\xi}J_0(\sigma_s)-\frac{\zeta}{\xi}-t_a\right]$$

as

$$C=\xi\left(\frac{\zeta}{\xi}+t_a\right)/[J_0(\sigma_s)-\mu J_1(\sigma_s)]$$

wherein $\mu=\sqrt{\xi}k/h_s$. This serves to complete the solution for $t(r)$ which finally reads

$$t=\frac{(1/\alpha_0+t_a)}{J_0(\sigma_s)-\mu J_1(\sigma_s)}J_0(\sigma)-\frac{1}{\alpha_0} \tag{8}$$

References

1. FORMULAS FOR CALCULATING THE TEMPERATURE DISTRIBUTION IN ELECTRICAL COILS OF GENERAL RECTANGULAR CROSS SECTION, T. J. Higgins. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 66, 1944, pp. 665–70.

2. THE NUMERICAL SOLUTION OF HEAT-CONDUCTION PROBLEMS, H. W. Emmons. *Transactions*, American Society of Mechanical Engineers, New York, N. Y., vol. 65, 1943, pp. 607–15.

3. THE PRANDTL MEMBRANE ANALOGY FOR TEMPERATURE FIELDS WITH PERMANENT HEAT SOURCES OR SINKS, P. J. Schneider. *Journal*, Institute of the Aeronautical Sciences, New York, N. Y., vol. 19, no. 9, 1952, pp. 644–45.

4. STEADY TEMPERATURE FIELDS IN ELECTRICAL COILS BY MEMBRANE ANALOGY, P. J. Schneider, A. B. Cambel. *Mechanical Engineering*, vol. 75, The American Society of Mechanical Engineers, New York, N. Y., Oct. 1953, p. 818.

5. RELAXATION METHODS IN ENGINEERING SCIENCE (book), R. V. Southwell. Oxford University Press, New York, N. Y., 1940.

6. HOCHSTTEMPERATUR STROMDURCHFLOSSENER SPULEN, W. Rogowski, V. Vieweg. *Archiv für Elektrotechnik*, Berlin, Germany, vol. 8, 1919–20, pp. 329–32.

7. INFLUENCE OF NONUNIFORM DEVELOPMENT OF HEAT UPON THE TEMPERATURE DISTRIBUTION IN ELECTRICAL COILS AND SIMILAR HEAT SOURCES OF SIMPLE FORM, M. Jacob. *Transactions*, The American Society of Mechanical Engineers, New York, N. Y., vol. 65, 1943, pp. 593–605.

◆

No Discussion

Production of Unipolar Air with Radium Isotopes

THOMAS L. MARTIN, JR.
ASSOCIATE MEMBER AIEE

Synopsis: This article is a general survey of recent development work on sources of unipolar air using alpha-emitting radioactive materials. Specific reference is made to a new device, operating at relatively high air speeds of about 450 feet per minute, which will provide an air output containing only positive or only negative ions. The method of controlling the density of this unipolar output is also treated.

UNIPOLAR air is air that contains ions of only a single polarity. There are many industrial applications in which it is necessary or desirable to have facilities for discharging electrified particles that adhere to materials or machines. It has also been suggested that atmospheric ionization can be used to cleanse rooms of contaminations such as smoke, dust, certain pollens, and some viruses. Such an action, in conjunction with air-conditioning systems, would prove extremely beneficial to sufferers of various allergies. With such an apparently diversified group of applications awaiting the availability of appropriate sources of unipolar air, it is not surprising that the development program has received considerable impetus.

Unipolar air may be produced in a number of different ways, but, for reasons discussed elsewhere, thermionic and alpha-particle radiation methods appear most suitable. It has been shown that thermionic methods lack the necessary degree of control, uniformity of output, and predictability desirable in ion sources.[1] Thus, it is concluded that ion production through alpha-particle radiation offers the greatest immediate area for potential development. This paper is concerned with recent efforts to produce unipolar air with alpha-emitting radioactive materials.

Choosing the Radioactive Emitter

In selecting the radioactive material for use in an ion generator, there are two main factors of importance to be considered:

Paper 53-351, recommended by the AIEE Basic Sciences Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Pacific General Meeting, Vancouver, B. C., September 1–4, 1953. Manuscript submitted May 13, 1953; made available for printing July 2, 1953.

THOMAS L. MARTIN, JR., is with the University of Arizona, Tucson, Ariz.

1. The half life of the emitter should be quite long. From the commercial viewpoint, a desirable half life would fall in the 3- to 5-year region, since this assures a reasonable shelf life for the manufactured ion sources. Moreover, this would be long enough to maintain a relatively constant ion output over the duration of most biological tests and would not require frequent replacement of units in commercial applications.
2. The emission of alpha particles should not be accompanied by strong beta or gamma radiation, nor should disintegration produce an isotope that will subsequently produce strong beta or gamma radiation.

With the foregoing factors in mind and with all cost considerations temporarily set aside, it was concluded that there are no satisfactory isotopes in the thorium or actinium series.

There are a number of alpha-emitting isotopes of radium that disintegrate with long half-life periods. However, radon is an ultimate product of the disintegrations of the heavier isotopes. Since it is a gas and radioactive, it is very dangerous to work with and should be avoided. Thus, all isotopes above radon in the series are ruled out.

Radium *G* is the nonradioactive end product of the radium series. Radium *F*, otherwise known as polonium, emits alpha particles only, reducing to radium *G* with a half life of about 140 days. It would be an ideal alpha-particle source if it had a longer half life. It is relatively inexpensive in comparison to other isotopes.

Radium *D* disintegrates to radium *E* and thence to polonium. In doing so, it radiates both beta and gamma rays of appreciable, but relatively low, energy. The transformation occurs with about a 22.2-year half life. An equilibrium mixture of radium *D* and polonium is thereby an effective alpha-particle source, having a half life of about 22 years.

The main disadvantage involved in the use of radium *D* is the presence of beta and gamma radiations. Fortunately the energy levels are comparatively low. Careful attention to shielding is required, but this is primarily a detail and not at all impossible to handle. Radium *D* is considerably more expensive than polonium. Nevertheless, in view of the 22-year half life, it could well prove to be more economical than polonium.

One more factor should be noted in connection with the use of polonium and radium *D*. It has been shown in the case of some types of polonium foils, that free polonium atoms are produced when alpha particles are emitted. In such cases, free polonium would be circulated into the air and would constitute a serious physiological problem. It appears that this problem has been overstated in some instances, and the advanced methods of foil preparation described in the next section have ended the problem.

Preparation of the Alpha Emitter

The alpha emitter used in all of the devices described in this paper was polonium foil. The method of preparing this foil is briefly summarized here.*

"The polonium is precipitated as a hydroxide mixed with silver powder. The intimately mixed powders are then introduced into a die and formed into a solid compact by heat and pressure. The radioactive component is then rolled to predetermined dimensions and is then ready for incorporation into a foil sandwich. The sandwich is composed of a base, consisting of a 1/8 inch silver block. A thin layer of inactive gold is laid over this base. A section cut from the radioactive compact is laid over this and another sheet of inactive gold is then placed over this. The entire unit is then compressed by heat and pressure into a solid fused block. The block is then rolled into ribbon form to an overall thickness of 0.005 to 0.006 inches. In this form, most of the foil thickness is due to the silver backing. Accordingly, alpha particles can be emitted only from the side without the silver backing.

"To add a harder and abrasion resistant surface, the rolled foil is given a final coating of about 0.4 micron thickness of rhodium or palladium. This improved foil is definitely *not* causing any of the old migration contamination troubles that are associated with polonium sources."

Important Aspects of Air Ionization with Alpha Particles

The main constituents of air are oxygen and nitrogen. These gases are diatomic, so that the discussion is mainly concerned with molecular rather than atomic effects. When alpha particles are emitted into this mixture, they collide with molecules, of both gases, producing free electrons and positive molecular ions of oxygen and nitrogen.

Oxygen readily forms negative molecular ions. Thus, except for a small number of electrons lost through recombination, the free electrons become attached to oxygen molecules. None of the electrons become attached to nitrogen molecules because it has been shown[2] that molecular nitrogen does not form negative ions. Thus, the ultimate result of the passage of alpha particles through air is the production of negative molecular ions of oxygen and positive molecular ions of nitrogen and oxygen, with the former predominating.

The mobility of an ion is its velocity in a unit electric field. If an absolutely pure gas could be obtained, at a constant pressure there would be a unique value for the mobility of an ion in that gas, starting about 1 microsecond after ion-formation.[3] However, such a gas does not exist and ion mobilities are found to vary over a wide range as a result of gas contamination and age of the ion.

As a result of the large variations in measured ion mobilities, there is no one value applicable for a given ion in a given gas. Thus, designers of ionizers should be wary of basing their design upon a special value for ion mobility taken from tables.

Losses

The number of ions that can be produced in air by an alpha emitter of known activity is easily computed. The number of ions obtained in the output from the ion tube can be determined experimentally. The ratio of the actual ion output to the theoretical maximum ion formation rate can be used as an index of the efficiency of the tube in utilizing the available ionizing capabilities of the alpha emitter.

This efficiency will always be less than 100 per cent. The reasons for this might be listed as follows:

1. Recombination losses;
2. Alpha-emission losses in the emitter;
3. Alpha absorption in the collimator;
4. Ionization losses in the separation process.

Each of these factors is discussed briefly in the following paragraphs.

Loeb[4] lists four types of ion recombination of potential importance here:

1. Preferential recombination;
2. Initial recombination;
3. Columnar recombination;
4. Volume recombination.

Regardless of the type of recombination, the rate of recombination can be computed from the general equation

$$\frac{dN(+)}{dt} = \frac{dN(-)}{dt} = -\alpha N(+)N(-)$$

where

α = recombination coefficient

---

* Permitted quotation from C. W. Wallhausen, United States Radium Corporation, New York, N. Y.

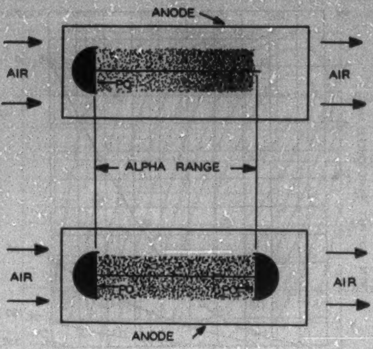


Fig. 1 (left). Two-electrode ion tubes

Desired ion pattern shown

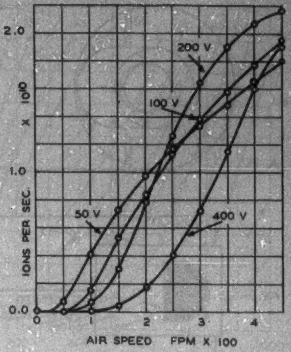


Fig. 3 (right). Characteristics of a 2-electrode ion tube; desired ion output versus air speed

$N(+)$ = density of positive ions

$N(-)$ = density of negative ions

However, the calculation is possible only if the recombination coefficient is known and, according to Loeb[5] "tables of values for α are more or less meaningless." Thus, recombination losses do occur, and the extent is probably appreciable, but it cannot be calculated with any degree of certainty.

Alpha losses in the emitter are largely caused by the construction of the polonium foil. From the foil preparation process described earlier, it is clear that on the average only half of the disintegrations will eject alpha particles into the air, because half of them will be absorbed in the silver backing. Alpha radiation parallel (or close to parallel) to the foil surface is weak because of the retardation of the rhodium or palladium coating. Also, because of this coating, all alpha particles will experience some retardation and loss of ionizing capability. Consequently, the effective alpha radiation from the foil is considerably less than the theoretical maximum of the polonium imbedded in the foil.

The alpha particles are emitted from the forward surface of the polonium foil in a fan shaped pattern.[6] This pattern is undesirable in some instances and collimating devices are used to confine the alpha radiation to a more beam-like pattern. This is accomplished by absorbing that radiation occurring at oblique angles from the foil surface. This absorption further reduces the effective activity of the emitter in producing ions.

To produce unipolar air from an original bipolar sample, some sort of ion separation mechanism must be used. It is inevitable that some of the desired ions will be lost in the process.

Ideally, these four losses could be separated and individually determined so that efficiency improvements could be made on an exact basis. Unfortunately, the author has not been able to do this and is not sure that it can be done.

Single Electrode Ion Tube

One of the first radioactive ion tubes produced in any volume to serve as a source of unipolar air was developed by J. C. Beckett. (The tube was described in detail in an unpublished paper by Mr. Beckett and H. H. Skilling.) It has proven to be a very useful tool in preliminary experimentation, and may represent the best solution to certain commercial requirements.

The tube consists of a single tubular electrode with a strip of polonium foil mounted on one side of this electrode. A rectified alternating voltage is applied between this electrode and ground. The dimensions of the tube, the location of the polonium, and the applied potential are all critically related to one another. The electrostatic flux lines go between the tubular electrode and nearby grounded objects such as the floor, walls, and ceiling of a room, or possibly even nearby people. The flux distribution in the tube opening is altered only slightly when the locations of these grounded objects are changed.

The operation of the tube is rather simply explained in principle. Alpha particles emitted by the polonium ionize the air inside the tube. If negative ions are desired in the output, the tubular electrode is made negative with respect to ground. The field configuration is such that negative ions drift out of the tube into the air while the positive ions are attracted to the electrode and are discharged. In principle, the tube provides a unipolar output.[7] It is a simple tube, easy to fabricate, and does not require the use of a blower fan to remove

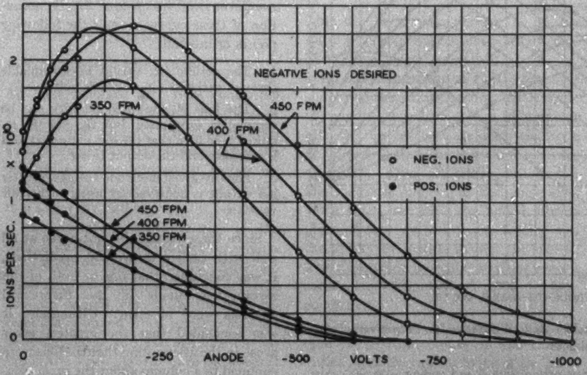


Fig. 2. Characteristics of a 2-electrode ion tube; ion output versus electrode voltage

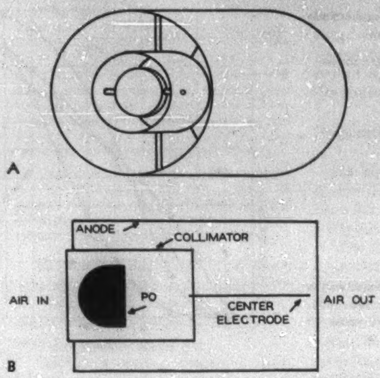


Fig. 4 (left). Triode ion tube

A. Assembly sketch
B. Cross-sectional schematic view

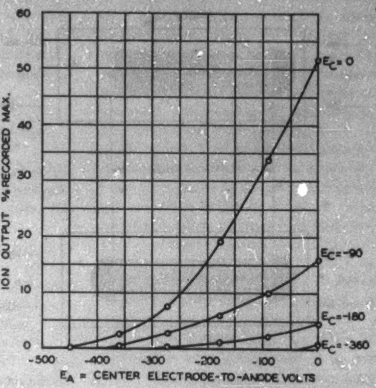


Fig. 6. Average characteristics of a triode ion tube; undesired positive ion output versus electrode voltages

the ions from the tube. This factor is of considerable importance in some areas of expected application. The main disadvantage of the tube is that it is not adapted to use at high air speeds, and would not be suitable in air-conditioning systems. A higher and controllable ion output would be advantageous in the same application.

Two-Electrode Ion Tubes

Two-electrode ion tubes were the first step taken in an effort to develop an ion source capable of operating at high air speeds. A cross-sectional schematic view of two early experimental models is shown in Fig. 1. There are two coaxial electrodes with the alpha emitting polonium facing axially down the tube. The ionization produced by the alpha particles would, desirably, be confined to a column about the innner electrode as shown in Fig. 1.

In operation, an electric field is established between the two electrodes and a relatively strong blast of air (450 feet per minute) is directed axially down the tube. The electric field strength varies inversely with the distance from the center electrode. Thus, the ions nearest this electrode are subjected to the strongest electrostatic forces. The ions of unwanted polarity are attracted toward the inner electrode into an ever increasing field intensity and are rapidly removed from the interelectrode space. The ions of desired polarity are forced toward the outer electrode into a field that becomes progressively weaker. These ions are thereby forced into the moving air stream and blown out of the tube before reaching the electrode.

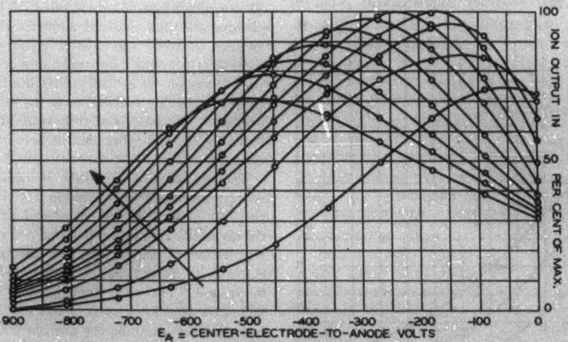


Fig. 5. Average characteristics of a triode ion tube; desired negative ion output versus electrode voltages

E_C is the parameter, changing in 90-volt steps commencing with 0 and ending with 900 volts. Direction of increase is shown by arrow. Air speed in Figs. 5–8 ≌450 feet per minute

Hypothetically, this tube could be made to operate as a true source of unipolar air. It can be shown that complete suppression of unwanted ions and complete removal of desired ions can be achieved for any combination of voltage and air speed if:

1. There is no fringing of the electric field;
2. The ions are confined to a cylindrical column, as shown.

Neither of these conditions exists, and this optimum kind of operation does not result.

Several of these tubes were constructed and tested. Some representative results are given in Figs. 2 and 3. Examination of these curves reveals the following points of interest:

1. The desired ion output is controllable through variations in electrode voltage.
2. The ions of unwanted polarity can be completely suppressed, but the voltage required is so high that the loss of desired ions is excessive and the output is low.
3. For any given electrode voltage, greater ion output is achieved by using higher air velocities.

Since these results did not agree at all well with earlier theoretical predictions, it was concluded that one or both of the assumptions made in the theoretical analysis were in error. Ultimately, it was concluded that the greatest error arose in the assumption that the ions were confined to a column about the inner electrode. As mentioned before, the emis-

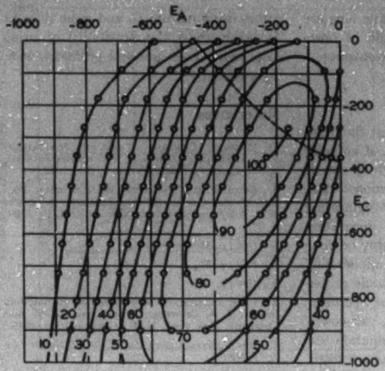


Fig. 7 (left). Average control characteristics of a triode ion tube; negative ions desired

E_A = center electrode-to-anode voltage

E_C = collimator-to-anode voltage

Dashed line is contour for zero positive ion output

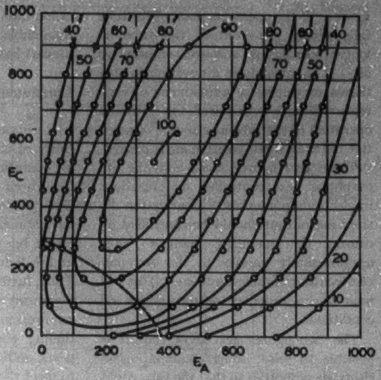


Fig. 8. Average control characteristics of a triode ion tube; positive ions desired

E_A = center electrode-to-anode voltage

E_C = collimator-to-anode voltage

Dashed line is contour for zero negative ion output

sion pattern from polonium foil is fan shaped and the ions are more or less uniformly distributed throughout the interelectrode space. It was concluded that improved operation would be obtained if the ion column could be confined more nearly to the assumed form. This required the introduction of a collimating device and led to the triode ion tube discussed in the remainder of this paper.

Triode Ion Tubes

Several different methods of collimating the alpha-particle radiation are being studied. However, only one type is covered here. The essential geometric configuration of the tube is shown in Fig. 4. The collimator is made of aluminum, and serves as the third electrode of the tube.

In taking data, the anode is used as the reference, or grounded, electrode. The tube voltages are then specified as follows:

E_A = center electrode-to-anode volts

E_C = collimator-to-anode volts

All tests reported here were conducted with an average indicated air speed of 450 feet per minute. Preliminary tests with other air speeds indicate higher ion outputs can be obtained at higher air speeds. This suggests that the separation losses are still appreciable.

The operating characteristics of the tube at an air speed of 450 feet per minute are shown in Figs. 5 and 6. Although these only show the case where negative ions are desired, essentially the same results are obtained when positive ions are desired. These are average characteristics in the sense that they are the average of several sets of data taken under varying degrees of air temperature and pollution. These two factors have some effect upon the results, but there were no variations in excess of 5 per cent, under the conditions tested.

The method of data presentation requires some explanation. Each point is expressed as a percentage of the maximum obtainable desired ion output under all electrode voltage possibilities. It was necessary to use some sort of normalizing procedure like this, since the continually diminishing activity of the polonium made the use of absolute values impractical. From these curves it is apparent that a wide latitude of control over ion output is possible through variations in electrode potentials.

In every case, the positive ions appeared to be more plentiful than negative ions. The average ratio of maximum values is approximately

$$\gamma = [N(+)\max]/[N(-)\max] = 1.15$$

This has been characteristic of all ion sources within the author's experience. The reason is not understood.

From a comparison of Figs. 5 and 6, it should be clear that the tube can be made to operate as a true source of unipolar air. Moreover, there are many electrode voltage combinations that allow unipolar operation. Thus, the triode ion tube fulfills the same function as the single electrode tube, but with the added advantages of easy control of ion density and the ability to operate at high air speeds. The 2-electrode ion tube was not even a satisfactory unipolar source under any conditions.

The maximum ion outputs obtained (the 100-per-cent points on the curves) varied with the radioactivity of the polonium. Although facilities were not available to measure the activity of the polonium, it appears that the maximum ion output with about 200 microcuries of polonium is about 1.2×10^{10} ions per second. This amount of radioactivity will produce a theoretical maximum of about 118×10^{10} ion pairs per second. Thus, the output is only about 1 per cent of the theoretical maximum. This is not a bad figure when corrected for alpha losses in the emitter. However, it could undoubtedly be increased somewhat through improved tube design and the use of higher air speeds.

Controlling the Ion Output

The method of controlling the density of ions in the unipolar output from the triode ion tube can be deduced from the average control characteristics given in Figs. 7 and 8.

In both of these figures, a potential plane has been constructed with the collimator voltage E_C plotted vertically and the center electrode voltage E_A plotted horizontally. Contours of constant ion output are drawn on that plane. That is, the 70-per-cent contour shows all of the possible electrode voltage combinations that will produce 70 per cent of the maximum obtainable ion output. Similar contours are drawn for all percentages from 10 to 100 per cent in 10 per cent increments.

The contour for zero undesired ion output is also shown on each of these figures. If the operating point of the tube falls within this contour the tube no longer operates as a source of unipolar air. It is particularly interesting to note that a more precise adjustment of electrode voltages is required for optimum operation of the tube as a source of negative air than when positive air is desired.

These two figures show how the ion output is best controlled by variations in electrode potentials. Clearly, the desired ion output should be controlled by causing the operating point to shift approximately parallel to the zero-per-cent contour of undesired ions. This will maintain a unipolar output under all conditions.

The two figures also show that it is possible to select some combination of electrode potentials so that the tube could provide positive or negative air just by reversing the applied potentials. Indeed, careful adjustment of these voltages will result in almost equal ion outputs in either case, since more positive ions are obtainable and the tube could be operated at less than 100-per-cent output.

Conclusions

It appears to the writer that the feasibility of controllable sources of unipolar air operating at relatively high air speeds has been conclusively demonstrated. The triode ion tube needs engineering development to reduce its separation losses, but this should preferably be done in the light of some specific application. Thus, further development should probably await the specification of operating requirements by biologists, air-conditioning engineers, decontamination experts, and so on, for no one tube is likely to meet the demands of all users.

The present development program will continue. Other designs are under consideration. Work projected in the immediate future will be concerned with tubes using radium *D*, and some thought is being given the possibility of using beta emitters in unusual geometric configurations.

References

1. Climate Control Through Ionization, T. L. Martin, Jr. *Journal of the Franklin Institute*, Philadelphia, Pa., vol. 254, no. 4, Oct. 1952.

2. Fundamental Processes of Electrical Discharge in Gases (book), Leonard B. Loeb. John Wiley and Sons, Inc., New York, N. Y., 1939, p. 289.

3. *Ibid.*, p. 33.

4. *Ibid.*, p. 37.

5. *Ibid.*, p. 147.

6. Ions, Electrons, and Ionizing Radiations (book), J. A. Crowther. Longmans, Green, and Company, New York, N. Y., eighth edition, 1949.

7. Ion Controllers. *United States Patent No. 2,579,399.*

8. *Ibid., United States Patent No. 2,589,613.*

9. *Ibid., United States Patent No. 2,594,777.*

◆

No Discussion

Gas Tube Filament Connections

MORRIS BRENNER
ASSOCIATE MEMBER AIEE

H. G. SCHMITT
NONMEMBER AIEE

THE OBJECT of this paper is to evaluate quantitatively the termination of inert gas-filled power tubes by contact pins and sockets and by flexible leads with pressure-type terminals.

Background

At first glance, the equipment designer who plans to incorporate a modern temperature-free gaseous thyratron or rectifier tube in a unit, should not expect any problem in providing adequate termination facility for such tubes. Unfortunately, such a seemingly simple design consideration may be the cause of much grief during the life of the equipment, if the tube is of the socket-base type. When gaseous thyratron and rectifier-type tubes were first introduced approximately 25 years ago, it was thought that sockets should be used in order to allow quick replacement. In making this decision, it was not realized at that time, that the use of a base and socket would jeopardize the reliability of the tube in service. Analysis of end of life tubes returned from the field over the past several years shows that over 9.0 per cent of such failures have been directly caused by inadequate socket conditions. Since tubes of this type are now one of the most important components in the power systems of modern industry, military aircraft, fire control equipment, and many other types of critical mechanisms, the importance of high tube reliability cannot be overemphasized.

Field reports indicate that a typical sequence of events in a socket failure is somewhat as follows: A new tube is placed in a socket and performs normally. After a short time oxidation or corrosion, aided by heat flow from the filament down into the base pin, increases the resistance at the filament pin contact. The temperature rises, accelerating the process until finally the increased contact resistance lowers the filament voltage enough to bring on a cathode failure after a relatively short life. Unsuspectingly a new tube is placed in the socket, and the process repeated but this time faster because of the initial oxidation of the socket clip and because the additional heat generated has caused the socket clip to relax some of its mechanical spring pressure. On some sockets, the variations in pin size and spacing within Radio and Television Manufacturers' Association (RTMA) tolerances are such that after a tube near the limits in one direction is used in the socket, the spring force will be seriously relaxed for a tube at the other limit.

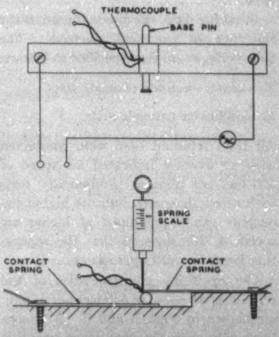


Fig. 1. Test setup for determining curves 1, 2, and 3 of Fig. 2. Contact spring force measured by looping a flexible cord around the top spring as near as possible to the pin. The force needed to open the contact was read on the spring scale

---

Paper 53-357, recommended by the AIEE Electronics Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Pacific General Meeting, Vancouver, B. C., September 1–4, 1953. Manuscript submitted June 4, 1953; made available for printing July 9, 1953.

Morris Brenner is with the Burndy Engineering Company, Inc., Norwalk, Conn., and H. G. Schmitt is with Electrons, Inc., Newark, N. J.

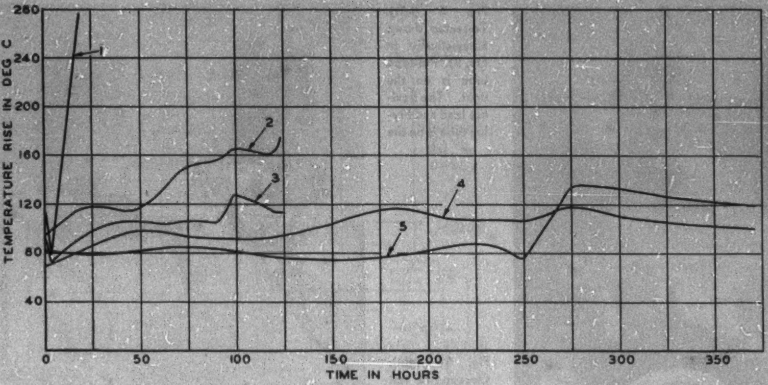


Fig. 2. Temperature rises of nickel-coated brass pins for different contact forces, determined from test setup shown in Figs. 1 and 3

Curve	Contact Force, Lb	A-C Amperes
1	1/2	30
2	1	30
3	2	30
4	2 1/2	50
5	2 1/2	50

Contact Requirements

The contact requirements for thyratron sockets are probably among the most severe to be found anywhere in industry, namely:

1. Fairly heavy currents must be handled: up to 26 amperes for the filament alone, plus load current as high as 77 amperes peak.
2. A fairly heavy wire must carry this current a short distance to a cathode running at 875 degrees centigrade; hence a substantial amount of heat is conducted to the contact.
3. The maximum total variation in filament voltage at the tube, for any substantial period without affecting tube life, is ±5 per cent. A change of contact resistance at each filament prong of 0.6 milliohms, generating a drop of 30 millivolts, would use up half of the allowable tolerance for low line voltage. In other words, it would affect tube life on many applications.
4. The contact must be maintained for several years, usually without any motion.

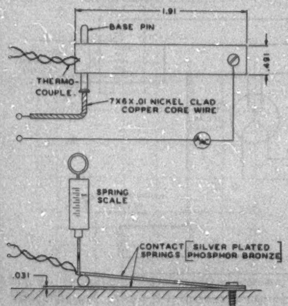

1.91
BASE PIN
.491
THERMO-COUPLE
7X6X.01 NICKEL CLAD COPPER CORE WIRE
A
SPRING SCALE
CONTACT SPRINGS
SILVER PLATED PHOSPHOR BRONZE
.031

Fig. 3. Test setup for determining curves 4 and 5 of Fig. 2. Procedure similar to one used for curves 1, 2, and 3

About 8 years ago, a 16-ampere tube was brought out on which the contact requirement was even more rigid: 34-ampere filament, and as high as 160 amperes continuously recurring peak load current. Realizing the pin-type limitations, this type was designed for bracket mounting with flexible leads ending in Hydent-type solderless-lug connections. The experience gained with this tube type was so gratifying that other types have followed, and the design has been adapted to many of the smaller 6-ampere diodes and thyratrons. Special thyratrons developed for military aircraft applications have incorporated the flexible lead connections with Hylugs for utmost dependability. Failures due to circuit connections on returned tubes of these types have dropped to an insignificant percentage. Out of many thousands of tubes now in service of this type only three instances of tube failure have come to light due to contact resistance in the filament lead. In each of these cases, it was evident that on installation one of the filament-lead nuts had only been set up with the fingers. Failure was prompt and obvious. The results have been sufficiently striking to warrant a more quantitative evaluation of the two terminating techniques.

Method

Spring Force

A base pin was arranged in a series circuit, Fig. 1, between two silver-plated phosphor-bronze spring clips from a typical tube socket. Contact forces were measured by looping a flexible cord around the top spring as near as possible to the pin. The mounting screws were adjusted to the desired force required to open the contact. This force was measured on a spring scale. Thirty amperes of alternating current were passed through the test setup at contact forces of 1/2, 1, and 2 pounds. Several sets of pins and springs were run simultaneously. Tem-

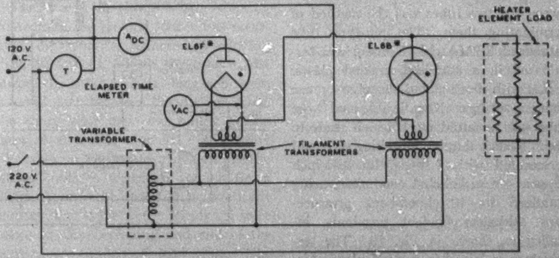


Fig. 4. Test circuit used for continuous load test

* Electrons, Inc., gas diode rectifier

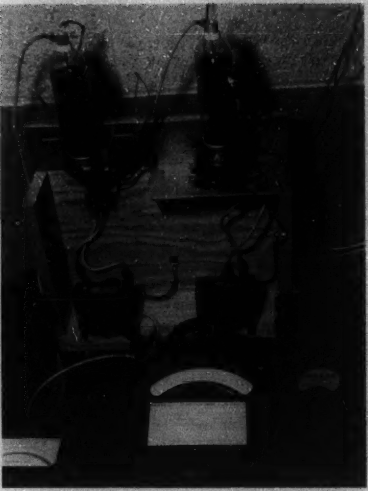


Fig. 5 (left). Test setup (shown schematically in Fig. 8). Pin-type tube is on the right. The flexible lead and Hy-lug tube is on the left

Fig. 6 (right). Closeup of test setup. Small openings in bottom of aluminum plate on the right are for voltmeter probes

peratures were measured by use of thermocouples welded to the end of the springs as shown. The results are shown on Fig. 2, curves 1, 2, and 3.

This test was repeated using the test setup shown in Fig. 3, except at 50 amperes, of alternating current (at a contact force of $2^1/_2$ pounds). The conductor attached to the base pin was used as part of the test circuit. The mounting screw in Fig. 3 was adjusted until the desired spring tension was achieved. The results are shown on Fig. 2, curves 4 and 5.

Tube Tests

Two 6-ampere xenon-filled half-wave rectifiers were set up in a test circuit, Fig. 4, and mounted as shown in Figs. 5 and 6, on 3/16-inch aluminum sheet, back and bottom. The only difference between the two tubes was the method of terminating them. The pin-type tube (standard RTMA *A4-81* base) was terminated in a standard molded plastic socket with silver-plated phosphor-bronze contact springs. The flexible-lead type tube was mounted on 1/4-inch studs to the 3/16-inch aluminum backing sheet and terminated with 7 by 6 by 0.01-inch copper-core nickel-clad conductor, using standard no. 10 closed-back pressure-type solderless Hydent terminals, installed as shown in Fig. 19. The terminals were joined to a small plastic terminal block with no. 10-32 brass studs. The leads from the terminal block to the filament transformer were no. 10 insulated flexible copper conductor (*AN10*). The leads from the socket of the pin type tube to the filament transformer were also no. 10 insulated flexible copper conductor (*AN10*).

The mounting method shown was decided upon in order to simulate as much as possible the way these tubes would actually be used in equipment. The pin-type tube is normally mounted in a molded plastic base which, in turn, is firmly mounted to a metal surface. On the other hand, the bracket type tube is usually mounted against a sheet-metal backing and the flexible leads are then brought out to the required connection points. The two tubes were mounted $7^3/_4$ inches apart on centers. The entire assembly was placed in a draft-free room, specifically used for tests of this nature. Tests were performed under continuous, Fig. 4, and on-off loads, Fig. 7.

Temperature rise and millivolt drop were used to evaluate the performance of the tubes. The temperature rise above ambient of all external tube parts from the anode cap to the filament connections

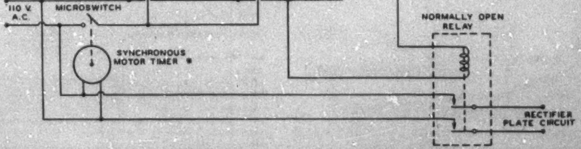


Fig. 7. Control circuit used for on-off load test

* 40 minutes on, 20 minutes off

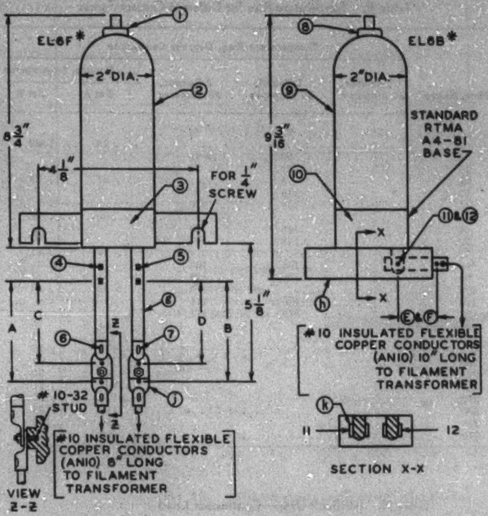

EL6F *
①
②
2"DIA.
8 3/4"
4 1/8"
FOR 1/4" SCREW
③
④
⑤
⑥
⑦
⑧
A
B
C
D
Z
Z
5 1/8"
(J)
# 10-32 STUD
#10 INSULATED FLEXIBLE COPPER CONDUCTORS (AN10) 8" LONG TO FILAMENT TRANSFORMER
VIEW Z-Z
EL6B *
⑧
⑨
⑩
2"DIA.
9 3/16"
STANDARD RTMA A4-81 BASE
⑪&⑫
X
X
(h)
(E)&(F)
#10 INSULATED FLEXIBLE COPPER CONDUCTORS (AN10) 10" LONG TO FILAMENT TRANSFORMER
(k)
11
12
SECTION X-X

Fig. 8. Diagram showing location of thermocouples during continuous and on-off load tests. (Numbers indicate actual thermocouple locations; letters indicate a-c millivolt-drop measurements)

(g) is 7 by 6 by 0.01 nickel-clad copper-core conductor, with silicone-impregnated Fiberglas sleeving

(h) is a standard 4-prong molded-plastic tube socket

(k) indicates the silver-plated phosphor-bronze contact springs in the tube socket

(J) identifies the tin-plated pure-copper compression-type terminals (Hylugs)

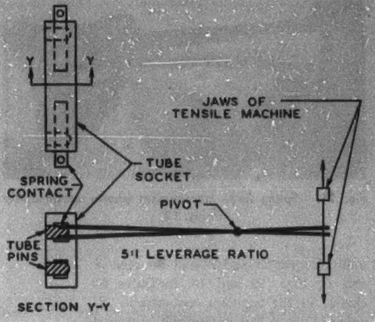

Y
Y
JAWS OF TENSILE MACHINE
TUBE SOCKET
SPRING CONTACT
PIVOT
TUBE PINS
5:1 LEVERAGE RATIO
SECTION Y-Y

Fig. 10. Schematic showing measurement of socket spring contact forces

Spring contacts separated until 0.002 shim could be inserted between contact and tube pin. Force reading on 100-pound indicator at this point is divided by 5 to get actual value of force required at contact

was measured by thermocouples. The location of the various thermocouples are shown in Fig. 8. Thermocouples 4 and 5 were located the same vertical distance from the bottom of the tube base as thermocouples 11 and 12 on the pin-type tube. Thermocouples were also placed on the pressure terminals. Voltage drops were measured across the filament connections of both types of tubes, Fig. 8.

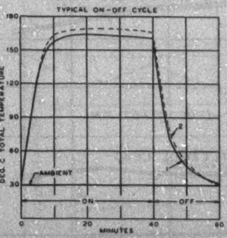


Fig. 9. Typical on-off cycle of 40 minutes on, 20 minutes off. Curve 1 is the anode cap of the flexible lead and Hylug tube. Curve 2 is the anode cap of the pin-type tube. There is sufficient off time for the hottest portion of the tube to return to ambient before the next cycle starts

Temperature and millivolt drop readings were taken on both tubes under continuous and on-off load conditions. The on-off cycle consisted of 40 minutes on and 20 minutes off. This was found to bring the parts of both tubes back to ambient temperature. Fig. 9 shows a typical heating cycle.

The force in the phosphor bronze spring contacts of the molded socket was measured by using a scissors-type device, Figs. 10 and 11. With the tube in the socket the prongs of the long arm were inserted just between the springs while the short lever arm was pulled in the jaws of a tensile machine. A reading was taken on the tensile machine when the spring contacts separated enough for a 0.002 shim to be inserted between the contact and tube pin. Readings were taken at the beginning and end of the continuous and on-off load tests. The filament transformer leads on the pin-type tube were soldered to the socket. The pressure terminals were screwed down to the no. 10-32 terminal strip studs with a torque of approximately 25 inch-pounds.

Results

Temperature Rise for Different Contact Forces

Fig. 2 indicates that low contact forces, in the order of 1/2 to 2 pounds for a phosphor-bronze silver-plated socket spring, will not result in a stable joint. At slightly higher forces, in the order of 2 1/2 pounds, the results are more or less stable, but stabilization occurs at relatively high temperature rises, in the order of 80 to 120 degrees centigrade.

Contact Force with Tube in Socket

Table I gives the results of measuring the contact force of the springs in the sockets with the tubes in place—before and after each test.

Terminal Force

Korn[1] gives the axial load in pounds per inch pound of applied torque for a no. 10 nut as 26 pounds. On the basis of an applied torque of approximately 25 inch-pounds, the total axial load on the pad of the terminal will be 25×26=650 pounds. This initial force is maintained because the small, constant, difference between

Table I. Contact Force

Load	At Beginning, Pounds	At End, Pounds
Continuous	7.3 to 7.8	3.86 to 3.85
On-Off	7.62 to 8.7	4.1 to 5.2

Fig. 11. Spring force measurement (shown schematically in Fig. 10)

millivolt drop *B* and *D* and between *A* and *C*, Figs. 12 and 13, indicates no change in the terminal connections. If either *A* or *B* had appreciably diverged from *C* or *D* during the progress of the test, it would have indicated a loss of contact force and an unstable contact condition.

Continuous Load-Millivolt Drop

The millivolt-drop readings of the continuous load test are shown in Fig. 12. One lead of the filament carries not only the filament current but also the additional plate load. The erratic behavior of the millivolt drop of the contact pins is the important thing to observe in Fig. 12. The voltage across both contact pins showed rising values, one more than the other. The drop across one pin increased by more than 100 per cent during the test.

Readings were taken from a point just below the tube base to the leads' own terminal, and to the filament transformer terminals. It should also be noted that the millivolt-drop reading for the flexible-lead tube was taken over a $5^1/_8$-inch dis-

Table II. Temperature Rise for Different Contact Forces

	Temperature Rise, Degrees Centigrade				
				$2^1/_2$ Pounds, 50 Amperes	
Time, Hours	$^1/_2$ Pounds, 30 Amperes	1 Pound, 30 Amperes	2 Pounds, 30 Amperes	Set A	Set B
1	117	97	84		
2				70	82
3	77	99	72		
19	176	11	85		
25		8		87	79
27		119			
44		117	110		
50				100	84
64		138	103		
72		148	105		
75				94	85
88		155	105		
99		165	128		
100				92	84
119		160	114		
123		175	114		
125				94	76
150				112	75
175				117	77
200				115	81
225				108	88
250				105	77
275				118	137
300				112	136
325				105	125
370				100	120

Table III. Millivolt Drop—Continuous Load

	Voltage Drop, Millivolts					
	Pin-Type Tube		Flexible-Lead and Hylug-Type Tube			
Time, Hours	Contact Pin E	Contact Pin F	A	B	C	D
0.4	55	22	27.3	50.5	27	49.5
3.9	57.5	24	26	50.4	25.7	50.0
24.2	60	21	25.1	47.3	24.8	46.5
49.8	52	26.5	24.4	48.7	24.0	48.3
118.6	52	29.2	24.6	50.9	24.0	50.6
141.6	55	30.5	25.5	51.7	24.7	51.3
191.7	56	29	24.2	52.9	23.5	51.9
219.2	58.5	36	26.2	54.3	25.8	54
284.1	57	44	26.2	53.3	25.8	52.8
308	60	41	27.3	52	26.8	51.6
356.8	60.5	46	25.6	49.2	25	48.8

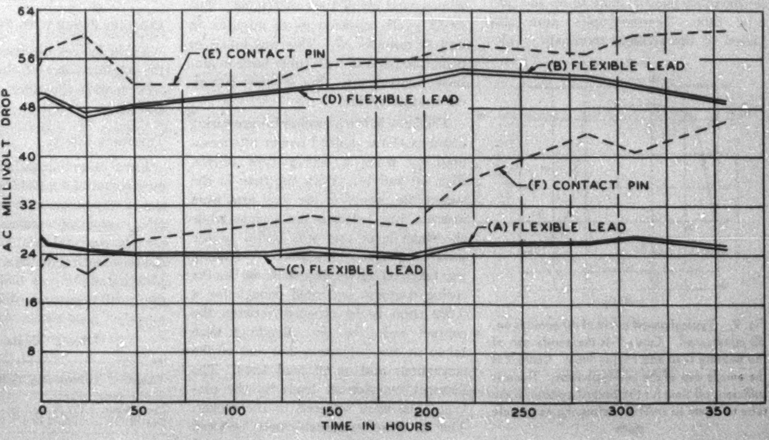


Fig. 12. Millivolt drops under continuous load for the pin-type tube, and flexible lead and Hylug-type tube (see Fig. 8 for points between which readings were taken). Solid lines are flexible-lead and Hylug-type tube, broken lines are pin-type tube in Figs. 12–15

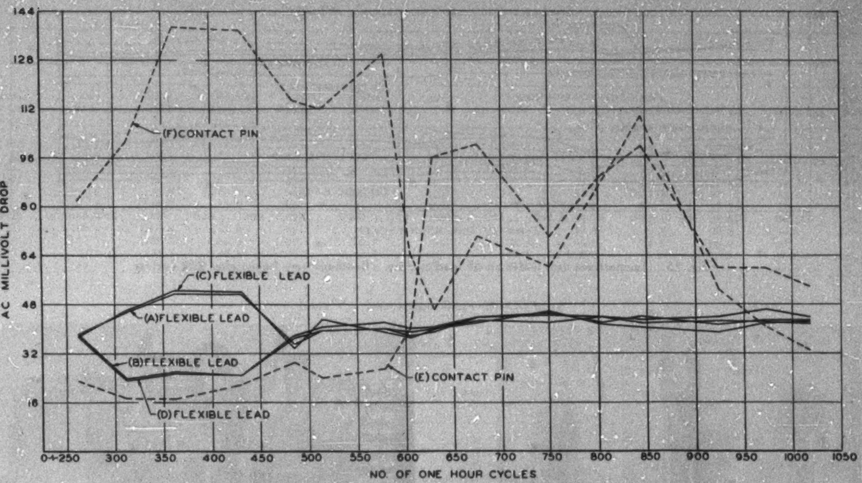


Fig. 13. Millivolt drops under on-off load cycling. Readings were begun after 262 cycles

tance, as compared to a distance of 1 inch on the socket-type tube. While the measuring distance was approximately 5 times greater on the flexible-lead type tube, the millivolt-drop reading was stable and lower compared to the contact-pin design.

Continuous Load—Temperature Rise

The temperature rise of the external tube parts are shown on Fig. 14. The anode caps and bulbs of both types of tubes stabilized at an approximately 140 degree-centigrade rise and remained there until the completion of the test at 360 hours. All the temperatures stabilized after approximately 50 hours. Continuing the test to 360 hours did not reveal anything which the first 50 hours had not already shown up. The outside of the phosphor-bronze springs in the molded socket had a 60-degree-centigrade rise, as compared to locations 4 and 5 on the flexible-lead tube, which stabilized at an approximately 55 degree-centigrade rise. The temperature rise of Hylugs 6 and 7 were considerably below the flexible-lead points at 4 and 5. The most marked temperature difference was between the tube base of the pin type which had a rise of 50 degrees centigrade, while the base of the flexible lead tube had a 25-degree-centigrade rise.

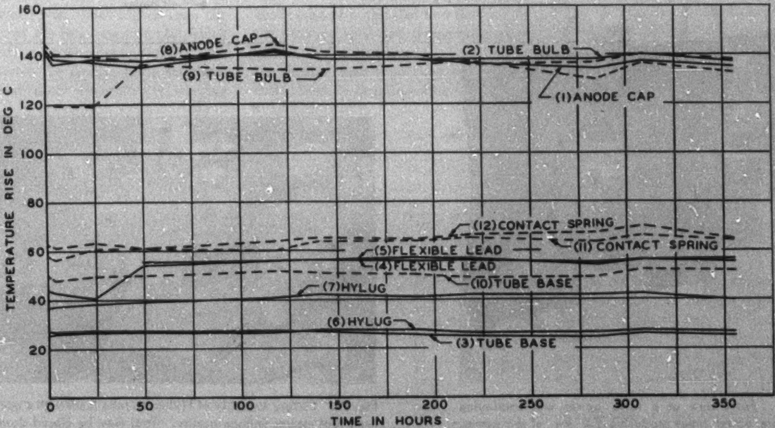


Fig. 14. Temperature rises under continuous load for the pin-type tube (see Fig. 8 for location of thermocouples)

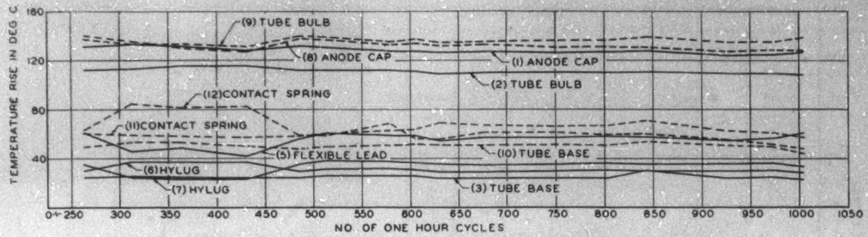


Fig. 15. Temperature rises under on-off load cycling. Readings were begun after 262 cycles

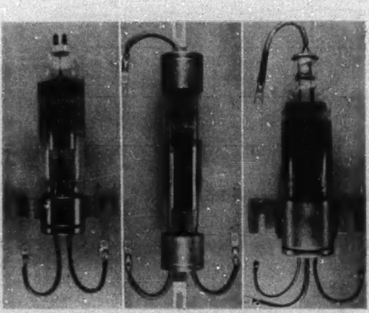


Fig. 16. Three types of gas-filled power tubes terminated with flexible leads and Hylugs. Note how the use of different types of mounting brackets plus flexible leads and Hylugs gives the equipment designer wide latitude in mounting these tubes in equipment

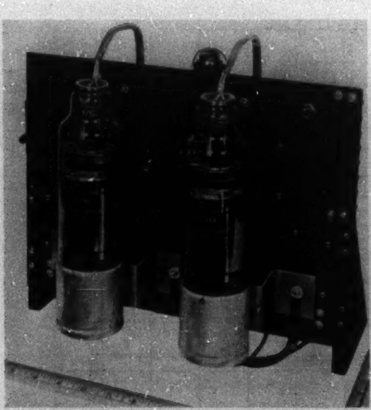


Fig. 17. Front view of a servo power unit employing 18-ampere power tubes producing 2.5 kw of d-c power. This unit weighs only 25 pounds and occupies approximately 1 cubic foot of space

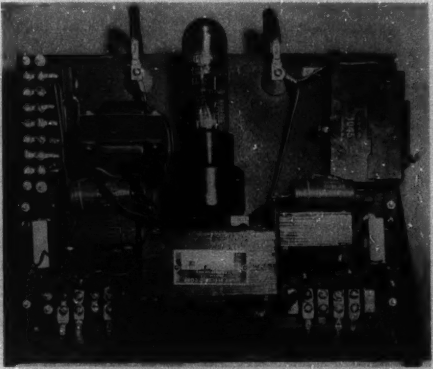


Fig. 18. Rear view of the servo unit described in Fig. 17. Note how the use of bracket mounting and flexible leads and Hylugs permits all the leads from the thyratrons mounted on the opposite side to be easily brought to the rear of the unit. This makes for a compact economical over-all design

Fig. 19. Factory assembly of Hylugs to tube leads with a ratchet-type hand tool assures uniform quality. Tool must be closed down all the way, thus assuring a perfect joint, before the operator can release the handle

Table IV. Temperature Rise—Continuous Load

Time, Hours	Temperature Rise, Degrees Centigrade											
	Pin-Type Tube					Flexible-Lead and Hylug-Type Tube						
	Anode Cap	Tube Bulb	Contact Spring 1	Contact Spring 2	Tube Base	Anode Cap	Tube Bulb	Flex Lead 1	Flex Lead 2	Hylug 1	Hylug 2	Tube Base
2	144.7		57.7	62.4	50	142	139.5		43.9	25.3	37	27.7
4	140.8	119.8	56.7	61.5	48.3	138.5	136.4		48	26	37.3	27.1
24	139.4	119.4	60.6	63.1	49.9	137.2	138.4		49.9	27	38.9	27.8
50	137.1	136.1	61.1	60.9	50.1	135.4	137.1	55.9	54.6	27.4	39.3	27.4
119	144.6	134.8	60.7	64.2	52.2	142.6	142	56.4	56.5	27	41.4	27.5
142	141.8	134.3	64.2	65.4	51	138.7	138.7	56.5	56.6	27.7	42.5	27
192	140.5	136	64.1	64.9	50.5	138.5	138.5	56	57.1	27.8	41.6	25.8
219	136.1	137.9	65.6	66.5	49.1	136.8	138.4	56.1	56.8	26.8	42.2	25.3
284	137.7	130.2	63.9	67.3	49.7	134.7	138.2	55	54.7	26.9	42	24.9
308	140.1	136.9	66.4	70.2	52.3	137.7	140.4	56.4	56.4	27.3	42.6	26.2
357	138	132.5	64.1	64.5	51.8	135	137.5	55.5	56.2	26.5	39.8	25.2

Table V. Temperature Rise—On-Off Load

No. of 1-Hour Cycles	Temperature Rise, Degrees Centigrade										
	Pin-Type Tube					Flexible-Lead and Hylug-Type Tube					
	Anode Cap	Tube Bulb	Contact Spring 1	Contact Spring 2	Tube Base	Anode Cap	Tube Bulb	Flexible Lead 1	Hylug 1	Hylug 2	Tube Base
262	140.7	137.7	60.7	63.2	49.7	131.7	112.7	61.1	35.4	30.3	24.7
312	136.4	133.7	59.5	85.0	53.7	133.6	113.8	46.1	24.7	37.7	25.4
362	130.8	134.4	58.9	82.2	53.6	132.6	116.3	49.3	23.8	37.8	24.7
432	127.8	131.9	57.7	83.1	50.8	128.9	116.4	42.5	24.0	37.5	24.9
485	138.6	140.8	59.0	59.5	48.6	132.6	112.8	56.8	35.9	30.0	24.6
514	137.7	139.7	61.1	59.5	50.5	131.2	112.9	61.3	38.9	32.1	26.8
578	133.5	136.0	63.5	69.3	51.1	128.1	112.5	59.5	38.2	32.3	26.5
605	134.5	138.0	69.1	56.5	52.1	128.0	111.5	58.9	37.0	31.0	26.0
630	132.3	135.2	69.5	56.1	51.4	126.7	109.2	54.7	35.7	29.8	24.2
675	133.5	135.6	67.9	61.6	51.2	127.7	110.6	57.5	34.9	29.5	24.2
749	131	136.5	66.6	61.5	51.5	126.5	110.5	57.3	35.3	29.7	24.0
802	131.7	136.7	66.8	59.8	50.8	127.2	110.2	58.5	36.6	30.7	24.2
844	131.1	139.6	71.2	60.3	53.9	128.6	111.1	57.9	35.6	29.6	30.2
923	127.9	135.4	62.9	55.7	50.9	124.7	109.7	55.1	35.7	29.9	23.9
972	128.8	135.4	61.3	52	49.8	125	109.6	56.8	36.3	30.3	24.8
1005	128.5	139	57.5	48	44	127	108.6	61.0	33.5	28.3	27.5

ON-OFF LOAD—MILLIVOLT DROP

The millivolt-drop readings of the on-off load test are shown in Fig. 13. As with the continuous load test, one contact pin of the pin type tube showed an erratic rising trend. It had millivolt drops over seven times higher than the drop across the other contact pin, which remained approximately stable. The drops across the flexible leads stabilized at approximately 40 millivolts, varying about this value during the course of the test.

ON-OFF LOAD—TEMPERATURE RISE

The temperature rises of the external tube parts are shown in Fig. 15. The anode caps of both tubes and the bulb of the pin-type tube had temperature rises of 135 degrees centigrade. The bulb of the flexible lead type tube was consistently about 25 degrees centigrade cooler. Similarly, the base of the flexible-lead-type tube was consistently 25 degrees centigrade or more cooler than the base of the pin-type tube.

The temperature rises of the contact springs and Hylugs show a direct relationship with the millivolt drop across these parts. When the voltage drop increased, the temperature rise also increased. When the voltage drop decreased, the temperature rise similarly decreased.

Table VI. Millivolt Drop—On-Off Load

No. of 1-Hour Cycles	Voltage Drop, Millivolts					
	Pin-Type Tube		Flexible-Lead and Hylug-Type Tube			
	Contact Pin E	Contact Pin F	A	B	C	D
262	23.0	82.0	38.2	38.2	37.6	37.6
312	17.5	101.0	45.6	23.7	46.0	23.3
362	17.2	139.0	51.6	26.0	52.6	25.6
432	22.0	138.0	51.7	24.6	50.2	24.6
485	29.2	115.0	35.0	37.9	33.8	37.2
514	24.0	112.0	39.5	40.5	43.0	39.5
578	27.0	130.0	40.0	42.0	39.0	40.0
605	40.0	65.0	37.5	40.0	37.0	39.0
630	96.0	46.0	39.0	40.5	39.8	40.0
675	100	70	43.5	43.5	42	42.5
749	70	60	45	44.5	45.5	42
802	90	88	42	44	41.5	44
844	100	110	44	43	40.5	42
923	60	53	41.5	44	39	43
972	60	41	43	46.5	42	42.5
1005	54	33	43	44	42.5	42

Conclusions

The results of both the continuous and on-off load tests show the flexible lead and Hylug tube to be significantly cooler than the pin-type design. With both tests, the tube bases of the flexible-lead and Hylug design were 50-per-cent cooler than the bases of the pin-type designs. During the on-off test, the bulb of the flexible-lead and Hylug design was 20-percent cooler than the bulb of the pin-type design.

The flexible-lead and Hylug-type design provides a more dependable, longer-lived tube than the pin-type design. The high stable contact forces achieved

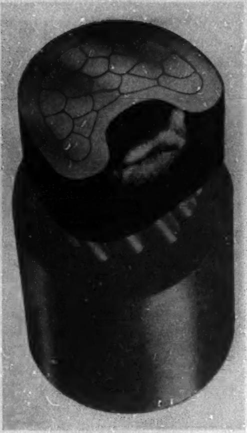


Fig. 20. Cross-section of a Hydent connection showing how the cup-shaped indentor cold-works the conductor. Each strand is compressed by the indentor into intimate contact with the Hylug barrel and the other strands, thus assuring a permanent stable joint

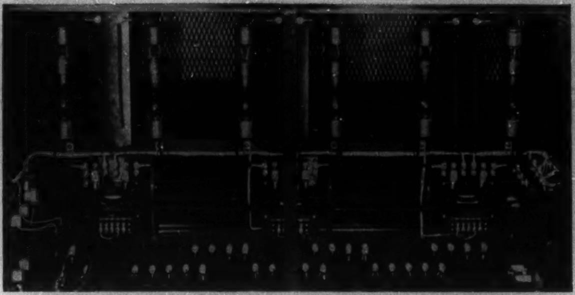


Fig. 21. Closeup of a heavy-duty power supply showing the use of six 16-ampere gas-filled rectifiers. The use of double-bracket mounted tubes and flexible leads with Hylugs gives the equipment designer great flexibility in mounting the tubes and provides better heat dissipation

by the flexible-lead-type tube results in greater voltage stability at the contacts. This means consistent operation within the tube's specifications, which results in more reliable and dependable tube operation. The allowable voltage tolerance of ±5 per cent is reserved for actual voltage fluctuations and is not used up almost immediately in the tube contacts. Instead of contact forces of seven pounds, which fall off to half that value in a few hundred hours of operation, contact forces of many hundreds of pounds are easily obtained and maintained.

The results indicate that the mounting bracket, flexible leads, and Hylugs drain away an appreciably greater amount of heat than the comparable tube of the pin and socket design. The importance of thermal engineering in the adequacy of electronic design[2] makes this a vital design consideration.

OTHER CONSIDERATIONS

Solder, with its attendant fluxes or acid, is completely eliminated with the flexible lead type tube. When the operating conditions of many tube applications are considered, this becomes extremely important. The high temperatures and corrosive conditions, coupled with the long life required, present a real quality-control problem to keep corrosion to a minimum.

Users at first were less sympathetic to the lug type than to the pin-and-socket type because of the longer time needed to change tubes, and because of the need of a tool to tighten the nut. Experience now indicates this should be considered as an advantage. Tests on tubes returned from service have traditionally disclosed a substantial percentage of tubes which pass all new tube tests. Investigating further, these tubes were usually replaced because the tube is the easiest thing to change. When any trouble occurs the first impulse is to replace all tubes and see if that fixes the trouble; if not, then a real search is initiated to run down the true cause. The Hylug seems to deter this philosophy, and results in a more careful analysis the first time.

There have been some disastrous epidemics of trouble with sockets. One was due to the substitution of another metal for silver-plated phosphor-bronze springs that had much better mechanical properties, but did not maintain good electrical contact. In another instance, spring-material quality control allowed a sizable percentage of certain sockets to relax their pressure with time. There have been socket designs that ultimately fill up with dirt, thus absorbing spring pressure and causing high resistance.

Design limitations make it difficult to provide a better factor of safety with a socket. For instance, if the contact pressure is increased, the tube cannot be inserted in the socket by hand unless longer travel and larger tube space is used. Since the required filament current for a given tube is fixed, unless a serious sacrifice in heating time can be tolerated, any higher filament voltage simply wastes heating watts in the tube and results in a larger tube. In view of these factors, a substantial improvement in contact method, such as flexible leads with Hylugs, is indicated.

In summary, experience and tests indicate that greatly improved unit reliability can be achieved with modern high-current temperature-free rectifiers and thyratrons, by eliminating base and socket construction through the use of bracket and pressure-lug construction.

References

1. Torque-Stress Load Relationship in Bolts and Screws, **A. H. Korn.** *Product Engineering,* New York, N. Y., Nov. 1943, pp. 752–54.

2. Temperature Prediction in Electronic Design, **P. J. Seigin, B. K. Hawes.** *Electrical Manufacturing,* Oct. 1952, pp. 116–20.

3. Tube Sockets for Industrial Use Must be Able to Take It. *Electronic Industries,* New York, N. Y., vol. 2, no. 6, June 1948, pp. 16–18.

4. Vacuum Tube Sockets for Industrial Equipment, **W. Clarkin, L. F. Biosca.** *Electrical Manufacturing,* New York, N. Y., vol. 48, no. 1, July 1951, pp. 72–77, 216.

5. End Cooling of Power Tube Filaments, **J. W. Clark, R. E. Neuber.** *Journal of Applied Physics,* New York, N. Y., vol. 21, Nov. 1950, pp. 1084–47.

6. Increasing Tube Reliability in Industrial Circuits, **D. G. Koch.** *Product Engineering,* New York, N. Y., vol. 23, no. 6, June 1952, pp. 175–181.

7. General Hydent Performance Test C4302–354. Burndy Engineering Company, Norwalk, Conn.

8. General Hydent Performance Test C4612–419. Burndy Engineering Company, Norwalk, Conn.

Discussion

Robert N. Eck and Paul M. Fischer (Cutler-Hammer, Inc., Milwaukee, Wis.): We are glad to have some scientific study given to this matter, since we have found tube sockets to be a serious source of trouble in filament circuits, particularly with the larger power tubes. We have known for a long time that reliable tubes, which should have given exceedingly long life, would fail after several thousand hours of life due to poor contact connections in the socket. On one particular make of socket, we found the spring clips to be nickel plated, which very much aggravated the oxide resistance difficulty. Excessive heat developed at the clip in the socket soon resulted in loss of tension in the phosphor-bronze clips. Obviously, the deterioration is rapid once it starts. To eliminate this difficulty, we have gone to the bracket-mounted tube wherever possible, even though it meant a tube of special design. In this way we have obtained positive connections and eliminated the possibility of tubes loosening in their mountings.

We feel that a further weakness in industrial sockets for power tubes is the inadequacy of the terminals. In most cases, it is very nearly impossible to attach the large lead from the filament transformer, to say nothing of the additional connections that sometimes need to be made at this same terminal. The panel mounted tubes, with insulated studs for connections, are much better from this standpoint. Also, the terminals are more accessible for maintenance and trouble shooting.

Some effort has been made in the direction of providing set screws to tighten the pin of the tube into the socket. This could be rough on the tube and a hazard in some ways, but it is a possible means for insuring better contact.

The paper emphasized the advantage of the terminal connection over the socket connection, thinking only of the base of the tube. While it is true that practically all of our troubles due to poor contacts originated in the base socket, we feel there is also a need for a flexible-lead connection for the anode. We believe that the elimination of the anode cap would be a big improvement. We have had the experience of having the anode cap fit so tightly that on removal, the seal between the anode terminal on the tube and the glass envelope of the tube would break. Hence, we recommend that the tube manufacturer provide tubes complete with anode, filament, and grid leads, together with the mounting bracket. It is only in this way that the tube user can fully realize the good performance designed into present day power rectifiers and thyratrons.

Railroad Communications Using High-Power Carrier Technique

A. K. HANSEN
NONMEMBER AIEE

W. W. H. DEAN
NONMEMBER AIEE

F. G. MULLINS, JR.
NONMEMBER AIEE

IN THE wilderness of the Quebec Labrador boundary a new Canadian industry is being established 360 miles from the nearest permanent settlement. This industry is the mining of iron ore and the location is Knob Lake, 360 miles almost due north of the small town of Seven Islands on the north shore of the Gulf of St. Lawrence.

It has been known for the past half century that large iron ore deposits existed in this area. Their location was so inaccessible, however, that the mining of it was commercially unattractive. The growth of air transportation during World War II made the project more feasible and the accelerated depletion of existing ore reserves made it desirable.

As a result of these factors the Iron Ore Company of Canada was formed in 1949. Two hundred million dollars capital was raised for establishing the necessary facilities to enable 70,000 tons of ore a day to be delivered to the docks at Seven Islands whence it can be transported by boat to steel mills throughout eastern North America and Europe.

A major part of the program is the construction of a railway over the 360 miles between Knob Lake and Seven Islands—a railroad that requires the heaviest rails rolled in Canada (132 pounds) and whose trains will carry the heaviest loads in North America (10,000 net long tons pulled by 4-unit diesel-electric locomotives, making a total of 14,000 trailing tons behind the locomotives).

Obviously such an extensive project must have excellent communications, not only for railway operation but in connection with the mining of iron ore and to provide public communications for the new town at Knob Lake. The responsibility of providing communications was assumed by the Quebec North Shore and Labrador Railway Company, a subsidiary of the Iron Ore Company of Canada. No major railroad has been built in North America in the last 50 years, so an exhaustive examination was necessary of the various communication facilities now available, whether conventional to railroad use or not.

Development of Specifications

Initial Considerations

The first thought was that space radio would be the ideal solution, as only three points required communications—Knob Lake and Seven Islands, and Oreway, the half-way point. The distance and traffic density, however, ruled out direct point-to-point radio on high frequencies since there are not enough high frequency channels available. Microwave or very high frequency (VHF) radio would be required with the necessity for a number of repeater stations located, on the average, not more than 30 miles apart.

Microwave began to look attractive with the addition of another requirement, that of signalling for the railroad operation. The railway line is single-track with passing tracks located, on the average, every 15 miles. It was decided to control all train movements from a central dispatch point using a centralized traffic control (CTC) system. In this system d-c signal pulses must be transmitted and received from each siding. It would therefore be necessary to place a radio station at each siding which could handle the CTC signals and which could also relay through traffic.

Two major requirements were necessary for such a radio system; a reliable source of power to operate the equipment, and over-all reliability. The great number of electronic tubes which would effectively be in series from end to end of the circuit (about 1,000) was of concern. First there is very little operating experience available on long unattended radio relay systems. With an average of 10,000 hours' tube life, two to three failures a day could be expected. Using duplicate equipment along with a preventive maintenance schedule and stand-by service personnel, it was expected that good continuity of service could be maintained during the time the railway was operating, but at high cost. During the winter, when the railway was closed down, the problem of maintenance would be more severe as many relay points would be accessible only by helicopter.

Paper 53-342, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Pacific General Meeting, Vancouver, B. C., Canada, September 1-4, 1953. Manuscript submitted June 8, 1953; made available for printing July 7, 1953.

A. K. Hansen is with the Quebec North Shore and Labrador Railway, Montreal, Quebec, Canada, W. W. H. Dean is with the Canadian Westinghouse Company, Ltd., Hamilton, Ontario, Canada, and F. G. Mullins, Jr., is with the Westinghouse Electric Corporation, Baltimore, Md.

The deciding point against radio, however, was the problem of supplying power. The cost of local generation and the problem of unattended operation at the many relay points was prohibitive. But power was also required at each siding to keep the batteries charged which are required for operation of the signal system. As a result of these considerations, it was decided to construct a power line along the right of way. Power for radio operation would then be available. But now the picture was altered. With a wire line available, radio was no longer required. Power line carrier could be employed eliminating the many repeaters and consequent cascading of tubes.

However, there was still the problem of reaching the sidings with CTC signals. A separate power line carrier channel could be run to each one from the control point, but the number of carrier channels is limited. It is also preferable to have direct connections in a CTC system wherever possible to ensure maximum reliability (although carrier equipment is used for this purpose). As the CTC operates on a direct current of less than 200 volts with shunt-keying across the line to provide signalling, it is essential that the line must have low series resistance and high shunt resistance.

The power line was to be constructed as a single-phase 23-kv balanced line using no. 4 steel-reinforced-aluminum-cable smooth-body 150-per-cent conductors. The same construction was suitable for a CTC line; thus a second line was strung on the same poles 13 feet below the first. As this lower line was isolated from the power line and was constructed with a barreling transposition every fifth pole, it should be possible to obtain a better signal-to-noise ratio by using it for the carrier. It could also be used for telephone signals from instruments located at each siding, and also from portable telephones used by train and work crews

The Final Specification

Thus the pattern for communications developed, and final specifications were written for the system. They call for:

1. Nine trunk voice communication channels between Seven Islands and Knob Lake, four to be accessible at Oreway, the remainder to be direct.
2. One direct-voice channel for dispatch use connecting Seven Islands, Oreway, and Knob Lake.
3. One spare voice channel for dispatch operations.
4. Six teletype channels between Seven Islands and Knob Lake, four direct and two with party line taps at Oreway.
5. Voice communication, using the physical line, from each end of every siding to the dispatch center.
6. Portable telephones capable of being tapped on the line at any point for communication with the dispatcher.
7. Automatic dial telephone installation to serve the town of Knob Lake and Oreway.
8. A switchboard at Seven Islands to control all traffic over the trunk channels to Oreway and Knob Lake. All channels terminating on this switchboard must meet long-distance telephone standards, regarding frequency response and signal-to-noise ratio, as they will be connected into the nation-wide long-distance system through the local telephone company at Seven Islands.

All connections to the communications line must be designed so as to protect personnel and equipment from any accidental application of the power line voltage, and from lightning.

Meeting the Specifications

Choice of Carrier Equipment

To meet the performance specifications, consideration was first given to the trunk (or carrier) channels. While not actually on a power line, the line construction and its proximity to the power-carrying pair required that power line carrier techniques be employed.

All power line carrier in Canada is licensed by the same Department of Transport authority that controls space radio. They advised that under no circumstances could frequencies above 200 kc be used because of possible interference in the air navigational band. As the railway straddles the main transatlantic airway from Montreal via Goose Bay, the importance of avoiding interference is obvious.

Having had the upper limit of transmission thus fixed, and the lower limit fixed at about 40 kc by normal power line practice, there was available 160 kc of frequency spectrum to fit in the required channels. Remembering that each channel must be duplex, it was necessary to locate 22 single-voice carriers and 12 single-teletype carriers in this space.

By making use of single side-band techniques it is possible to space carriers 5 kc apart, using up 110 kc of spectrum for the voice channels. As only six teletype channels are required in each direction, frequency shift type carriers spaced 0.8 kc are supplied. This uses up another 10 kc, approximately, of spectrum, or no more than a duplex-voice channel incorporating voice frequency telegraph.

The total carrier requirements could be provided for in 120 kc of spectrum space. The remaining 40 kc is then left available for additional channels, guard bands and co-ordination with low-frequency radio services in case of mutual interference.

Fig. 1. Typical pole line construction

To meet the required transmission standards, it is not possible to transmit a carrier the entire 360 miles. Also, some channels must be accessible at Oreway, the halfway point in the line. All, therefore are repeated at this point. From the discussion on frequency requirements it may be seen that there are no frequencies available to allow a second set to be used beyond the relay point. All frequencies in the second section of the line are therefore duplicates of the first, but arranged so that each channel is received at or transmitted from Oreway on the same frequency in each direction. In this way the power level, either transmitting or receiving, is the same each side of Oreway on the duplicate frequencies. Good isolation of the equipment in the repeater installation by shielding, separate ground lead, etc., minimizes coupling between equipments. Also in one direction the physical line is replaced by a coaxial cable for 500 feet, to reduce mutual coupling between the two line sections. Detailed information on the carrier equipment and its application appears in the section "Equipment Application to the System."

Telephone and Associated Equipment

Associated with the carrier equipment is terminal telephone and teletype equipment. At Knob Lake a 300-line, completely automatic dial-type exchange is being installed. It will serve the various offices connected with the mining operation as well as the town site, even providing pay phones. The telephone installation is made in such a way that it may be expanded in the future to match town growth.

A similar but smaller dial exchange will serve the dispatch center at Oreway. In addition to the regular telephone system, the dispatchers have their own private carrier channel with a loudspeaker monitor connected across it. They can thus hear all incoming calls, answering with a standard telephone handset, which cuts off the loudspeaker as soon as it is lifted off its cradle.

The Seven Islands installation consists of a switchboard on which all the carrier channels are terminated. Connections are made here to the company's private branch exchange telephones, to the local Seven Islands telephone system, and to long distance through the Seven Islands town operator.

The teletype instruments are all page-type printers and are connected to provide local copy. That is, all messages print on the local machine as well as the remote one. The teletypes are connected into the carrier system through the telephone cables, but have no other connection with the telephone equipment.

For communication from along the right-of-way, booth-type telephones are permanently installed at each siding. In addition, portable instruments are supplied and provided with hook sticks to enable temporary connections to be made to the line at any point by train and work crews.

System Operation

This system is designed to provide the most flexible possible operation while retaining control through the Seven Islands switchboard. The operator at this point can select carrier trunks connected to either Oreway or Knob Lake telephone equipment, and dial the desired local number. At the same time, subscribers at either of these points, by dialing "operator" will automatically pick up the first idle carrier circuit and signal the operator. For convenience, and as heavy traffic is not anticipated, it is possible to dial directly between Knob Lake and Midway by prefixing the desired number with "9."

Instead of leaving the spare dispatch channel idle except when required as a result of failure of the regular channel, it is connected to the switchboard and Knob Lake exchange. By making it the top level on the trunk selectors, it is the last channel to be used in regular telephone service, and consequently is the most likely one to be available. This spare is also connected to the dispatchers' locations in such a way that, by operating a switch, the dispatch equipment is connected to it and the exchange equipment disconnected, whether it is in use or not. The railway dispatchers are therefore, given complete priority.

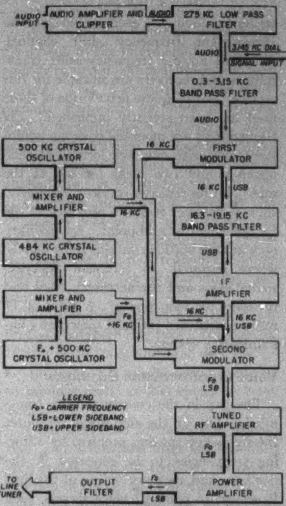


Fig. 2. Block diagram of single-side-band transmitter

Wayside Protectors

The terminal carrier equipment is connected to the communication line with conventional high-voltage capacitors just as if the line were energized. For the wayside and portable telephones, special protectors are required. A unit is available, known as the Osborne telephone protector, which is intended for coupling telephones to lines which are exposed to the application of high voltage.

This device consists of a high-voltage isolation transformer with the primary center tapped and grounded. There is also a double-coil relay series connected on the line side. Fault current, either between lines or to ground, causes the relay to close and ground both sides of the line. This relay has carbon contacts and is constructed so that quite heavy fault currents may be passed to ground without causing damage.

For this particular application two additional features are required. First, except under fault conditions, there must be no d-c path between lines or to ground. Second, and very important, a very small bridging loss must be caused at carrier frequencies, when the device is connected across the line. Direct current may be blocked by installing capacitors in series with each line just ahead of the center tapped transformer. Because of the practical limitation on capacity, however, these effectively prevent the flow of 60-cycle power current in case of a fault. This is overcome by placing a gas-type discharge tube from each line to ground. Finally a low-pass filter is added which will pass voice and direct current but which offers a very high impedance at carrier frequencies.

The final device, all housed in a cast aluminum enclosure, consists of the following components, starting at the line terminals:

1. Hook operated disconnect switch.
2. Grounding relay.
3. Discharge tubes, 400 to 600-volt rating.
4. Low-pass filter.
5. Blocking capacitors, 2 microfarads.
6. Isolation transformer.

CTC signal connections are tapped off ahead of the blocking capacitors and voice signals from the secondary of the transformer. (The CTC equipment contains filters to prevent it from interfering with voice signals.)

It will be noted that the discharge tubes have a 400 to 600-volt rating, with a nominal breakdown at 450 volts. This high voltage rating is necessary as the peak carrier voltage plus d-c CTC signals can theoretically exceed even 900 volts between lines. This peak would be reached at such infrequent intervals that any clipping caused by the tubes discharging is quite acceptable.

In this circuit the grounding relay and discharge tubes are placed ahead of the filter to obtain the maximum protection. The filter impedance at carrier frequencies is lowered by doing this; nevertheless, the loss introduced by connecting the protector across the line is less than 0.2 decibel at all carrier frequencies. The transmission loss through the device from the telephone instrument is about 2 decibels, while the d-c resistance for CTC signals is less than 2 ohms per leg.

Obviously, the portable telephones could not be supplied with such an elaborate device as the device is about three times the volume of the telephone unit itself. Since the portable units are connected only temporarily and for short periods, it was considered permissible to allow a slightly greater bridging loss. Also it has been observed that users very seldom ground portable units even when facilities are provided, and in the area where these sets are used, only elaborate and deep ground systems are effective because of soil conditions.

It was decided, therefore, to build as much insulation as possible into the portable telephones, to provide isolation for direct current, and to provide the highest impedance possible in the available space for carrier frequencies.

This was accomplished by using an isolation transformer similar to that in the fixed protector, but having a split primary winding with a blocking condenser in series to provide d-c isolation. This transformer is followed by a small, telephone-type carrier filter. An interesting feature of the isolating transformer is that, to save space and weight, yet provide good insulation, the transformer and its blocking capacitor are molded in a single unit. The thermosetting plastic used produces a unit which is capable of passing *MIL-T-27* specifications for class 1, grade A service.

Equipment Application to the System

EQUIPMENT SPECIFICATIONS

The preceding indicates in a general way the equipment chosen to meet the system requirements. A more detailed description of the carrier equipment will now be given. In particular, the voice carrier is described, as this equipment must operate through lines having an attenuation of 50 decibels, incorporate a repeater, and provide a signal to interference ratio and frequency response comparable to toll telephone service.

TELETYPEWRITER CHANNELS

Consideration was initially given to the use of a carrier channel, modulated by tones, to provide the teletypewriter service. In fact, it was contemplated that a 2-frequency duplex carrier could be modulated with six tones for the teletypewriter channels and ten tones for dial-signalling on the automatic telephone channels. This arrangement provides an economy in both equipment and frequency spectrum. Its disadvantage is, obviously, the fact that the entire communication system is dependent upon a single carrier channel.

This disadvantage could be offset to some degree by a spare carrier channel, manually or automatically switched into use if the primary channel should fail. However, in the final system, teletypewriter functions are each placed upon an independent frequency shift channel. Standard a-c/d-c frequency shift transmitters and receivers are suitable for this service. An adequate description of those used has been recently published.[1] Six transmitters are mounted upon an equipment rack with one rack located at each terminal and two racks located at the mid-point for repeater service. Similarly, six receivers per rack are provided at the terminals and at mid-point. An auxiliary rack at each station contains relay panels, where required, to establish teletypewriter loops and the necessary fuses and rectifiers for loop power. A total of 11 racks constitutes the teletypewriter communication system.

VOICE COMMUNICATION CHANNELS

The gap in the communication transmission line at the mid-point alleviates the frequency spacing problem by allowing frequencies to be repeated in the northern and southern halves of the line. However, to place 22 voice signals, each with an audio response to 3,150 cycles, within the 40 to 200-kc range, in addition to the 12 previously described frequency shift signals, is thought to be a unique application in the power line carrier field. Single-side-band carrier operation makes this possible.

Use of single-side-band power line carrier transmission and reception in the United States and Canada is not new.[2,3] However, the use of selective filters for reduction of adjacent channel interference and spurious transmitter outputs[1,4] came into wide use in power line carrier practice after the development of the old, earlier type of single-side-band equipment.[3] Even with one side band eliminated, the closest usable spacing with this equipment was about 8 or 10 kc. To permit the closer channel spacing required for this job either modernization or complete redevelopment was necessary. The latter course was chosen, with 5-kc spacing between adjacent transmitters as the goal. New techniques, which were developed for frequency division multiplexing over microwave[5] were found to be readily adaptable to power line carrier.

SINGLE-SIDE-BAND TRANSMITTER

A block diagram of the single-side-band transmitter applied to this system is shown in Fig. 2. Crystals are temperature-controlled to meet Canadian Department of Transport stability requirements. High-quality crystals of the frequencies used are readily obtainable. The 500-kc and 484-kc crystals are permanently installed. Outputs of the oscillators, controlled by these crystals, are mixed and amplified, the 16-kc frequency difference being used in the first modulator, regardless of the desired carrier frequency.

The output of the 484-kc oscillator is used again in the second mixer, where, in combination with the third crystal oscillator output, a frequency F_0 (the desired carrier frequency) plus 16 kc is derived, and amplified to a suitable level for use in the second modulator. To obtain single-side-band operation through the use of dual modulators permits the shaping of the passband by fixed filters, rather than units which must be changed with each change in output frequency.

A conventional audio amplifier contains two gain-controls to adjust input and clipping levels separately. A low-pass filter, flat within 2 decibels to 2,750 cycles and 40 decibels down at 3,145 cycles (the frequency of the dial-signalling tone) is used, together with a band-pass filter, flat within 2 decibels from 300 to 3,150 cycles, at least 30 decibels down at 120 cycles and 4,000.

A balanced ring modulator of conventional type, utilizing germanium diodes, combines audio output from the filters and the 16-kc output from the first mixer amplifier to produce side bands 16.3-kc to 19.15-kc and 15.7-kc to 12.85-kc. The 16-kc component is removed by the modulator.

A second band-pass filter, flat within 3 decibels from 16.3-kc to 19.15-kc, down 34 decibels at 15.7-kc and 30 decibels at 21-kc, passes the upper side band. An intermediate frequency amplifier compensates for filter losses and provides a convenient point for reinserting the 16-kc signal as a pilot carrier.

A second modulator, similar to the first, provides the means for deriving the output frequencies. The 16-kc carrier, together with the 16.3-kc to 19.15-kc side band are mixed with the F_0+16-kc output of the second crystal oscillator mixer. Difference signals produced are F_0 as pilot carrier and a lower side band, F_0-(0.3-kc to 3.15-kc). Sum signals produced, F_0+32-kc pilot carrier and a corresponding upper side band, are rejected by the triple-tuned radio frequency amplifier. Staggered tuning is used for the amplifier circuits to provide an adequate passband for the desired signals. An untuned class *A*, push-pull, parallel power amplifier, using eight *25L6* tubes, is capable of providing 16 watts output into the filter.

Two parallel-tuned traps in the output filter attenuate the second and third harmonic frequency outputs, and a series resonant circuit, of which these traps are a

Table I. Transmitter Electrical Specifications

Supply voltage	Nominal supply voltage, 129 volts d-c; heater voltage, 129 volts a-c or d-c
Supply variation	±10 per cent
Operating frequency	40 to 200-kc
Frequency stability	±8 cycles from 0 to 60 degrees centigrade
Output power	Ten-watt capability at filter output
Harmonic output	60 decibels below transmitted output when terminated in 60-ohm resistive load
Audio input	Full modulation with −20 decibels referred to 1 milliwatt (dbm) input at 800 cycles

part, passes the output frequency to the coaxial cable connection to the line tuning device.

Specifications for the single-side-band transmitter are shown in Table I.

SINGLE-SIDE-BAND RECEIVER

Fig. 3 shows a block diagram of the single-side-band receiver.

The carrier input signal, F_0 pilot carrier and lower side band as produced by the transmitter, is applied to the receiver input filter. This filter is interchangeable at 5-kc intervals throughout the 40 to 200-kc frequency band. Characteristics of a typical filter may be illustrated for a frequency, F_0, of 100 kc. This filter is flat within 2 decibels from 96.85 kc to 99.7 kc. It is down 30 decibels at 94.7 kc and 101.85 kc, and is down 70 decibels at 91.55-kc and 105-kc.

This permits 2-frequency duplex operation with the receiver separated by 10 kc from the full power output of the transmitter. A radio frequency input control permits adjustment of input signal level.

The signal, after passing through the input filter is impressed on a clipper-cathode follower, the output of which is mixed in the first demodulator with the F_0+16-kc output of the crystal oscillator and amplifier. Any difficulty from frequency instability of the crystal oscillator is eliminated, since the error is applied in translation to the pilot carrier and side band equally.

Of the outputs produced in the first demodulator, only the difference frequencies are used, 16 kc for the pilot carrier, and 16.3 kc to 19.15 kc for an upper side band. Characteristics of the 16.3-kc to 19.15-kc band-pass filter are similar to those of the filter used in the transmitter. Three single-tuned intermediate frequency amplifier stages provide additional selectivity and the necessary gain for the receiver.

The 16-kc pilot carrier signal is extracted from the output of the intermediate frequency amplifier and is further amplified through 16-kc tuned circuits, rectified, and the resulting voltage applied back to the intermediate frequency amplifiers as conventional automatic gain control. This 16-kc output of the intermediate frequency amplifier is further used to synchronize a self-excited oscillator, which provides the final translating signal. The 16-kc pilot signal could be applied directly for translation, but the arrangement used allows for fades or other temporary loss of the pilot signal, the self-excited oscillator continuing to function near the desired frequency even in the absence of the synchronizing carrier.

The second demodulator, similar to the first, is a balanced ring type. It mixes the 16.3-kc to 19.15-kc side band with the 16-kc output of the synchronized oscillator to produce audio of 300 cycles to 3,150 cycles.

A low-pass filter for side-band separation, and a low-pass filter for signalling tone exclusion in the receiver, permit extraction of the 3,145-cycle dial signalling tone and allow elimination of spurious outputs above and below the desired 300 to 2,750 cycles at the output of the audio amplifier. Specifications for the single-side-band receiver are shown in Table II.

Table II. Receiver Electrical Specifications

Supply voltage	Nominal supply voltage, 129 volts d-c; heater voltage 129 volts a-c or d-c
Supply variation	±10 per cent
Operating frequency	40 to 200-kc
Temperature limits	0 to 60 degrees centigrade
Sensitivity	One-watt pilot carrier and 4.5-watt peak power side band produce ±8 dbm audio output through a 70-decibel (db) noiseless channel
Selectivity	0.3-kc and 3.15-kc below carrier frequency, 3 db At ±3.575-kc from channel center frequency, 80 db
Automatic gain	Less than 6-db increase in audio for 40-db increase in radio frequency

SINGLE-SIDE-BAND ASSEMBLIES

A single-side-band transmitter and receiver, together with the necessary rectifiers for a-c operation and other auxiliaries, constitute a voice communication assembly. Three different types are included. For the channel designated for the railroad dispatcher, 2-wire telephone termination and amplifiers, and speakers for voice calling are provided. For the equipments terminating in automatic telephone switchboards, 2-wire telephone and a 3,145-cycle frequency shift tone for dial signalling are provided. For the equipments used for repeaters at the mid-point, 4-wire audio termination is used, the receiver audio output from one direction being connected to the corresponding transmitter audio input in the opposite direction. Signalling equipment and hybrids are eliminated in the latter. Equalizer amplifiers are added on the automatic telephone channels to provide the desired flatness in the audio response. A total of 44 racks constitute the carrier voice communication equipment, 11 racks at the south and north terminals, and 22 racks at the mid-point.

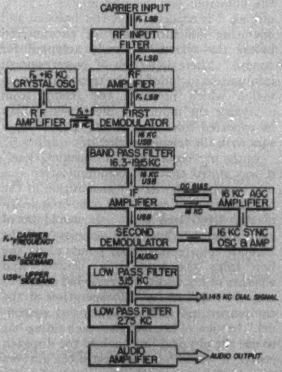


Fig. 3. Block diagram of single-side-band receiver

COUPLING PROBLEMS

Two problems were outstanding in this application, both relating to coupling. First, operating all transmitters in parallel, with the requirement that spurious signals be 30 decibels or more below desired signals in the receivers, required that intermodulation products in the receiver passbands be 80 or 90 decibels below the transmitted output levels. The second problem related to broad-band coupling to the line.

Transmitter output filters provide some isolation, although it is anticipated that transmitter outputs may have to be reduced to prevent the transmitters from overloading each other. All transmitters at one terminal are grouped, with 5-kc frequency spacing, into one end of the frequency band with a high-pass or low-pass filter as required to protect the receivers, which are grouped in the opposite end of the frequency band. All frequency shift channels are grouped near the center of the frequency band with a band-pass filter at the frequency shift transmitter outputs to prevent the generation of voice intermodulation products

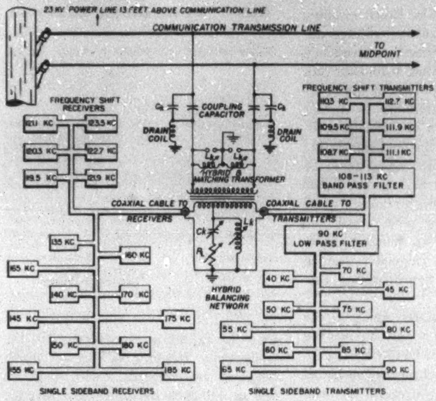


Fig. 4. Coupling method used for paralleling transmitters and isolating receivers at terminal stations

in the frequency shift transmitter outputs. One additional provision had been suggested, that of using a hybrid to separate transmitters from receivers.

A subsidiary problem was presented with the discovery that the available hybrid transformers produced intermodulation products beyond the tolerable limits. A special transformer was designed, utilizing a Hypernik toroidal core and a very low magnetizing current. To avoid the necessity of designing a separate matching transformer, the hybrid and matching functions were combined.

For true broad-band coupling, a constant-*k* high-pass filter is suitable, provided the cut-off frequency can be set sufficiently low to allow impedance matching at the lowest frequency used. A coupling capacitance of 0.005 microfarads, such as is available for higher voltages, results in a cut-off frequency of 98-kc for a 325-ohm termination. Fortunately, this application involves a low-voltage circuit of 11.5-kc, line-to-ground, and 0.02-microfarad capacitors having a cut-off frequency of 24.5-kc can be economically used.

Fig. 4 shows how the necessary coupling elements are combined. A combined hybrid and matching transformer is situated with constant-*k* filter elements in its primary and secondary. The constant-*k* element in the primary, together with the balancing resistor in the primary, constitutes the balancing element of the hybrid. The input to the coupler matches the 60-ohm impedance of the normal coaxial line, while the coupler output is adjusted to match the impedance of the communication line.

CONCLUSION

The elements discussed as constituting this system, that is, frequency-shift transmitters and receivers for the teletypewriter channels; single-side-band transmitters and receivers for voice communication with automatic dial telephone switchboards; and broad-band coupling to a communication line in joint use with direct current for centralized railway traffic control and audio for linkage of base railroad radio communication circuits, provide a unique communication system of great diversification, which involves several novel approaches to the railway communication problem.

References

1. A Flexible New Line of Power-Line Carrier Equipment, **Frank B. Gunter.** *AIEE Transactions*, vol. 72, pt. III, June 1953, pp. 504–16.

2. A Comparison of Amplitude-Modulation, Frequency-Modulation and Single-Side-Band Systems for Power-Line Carrier Transmission, **R. C. Cheek.** *AIEE Transactions (Electrical Engineering)*, vol. 64, May 1945, pp. 215–20.

3. A New Single-Side-Band Carrier System, **B. E. Lenehan.** *Electrical Engineering*, vol. 66, June 1947, pp. 549–52.

4. A Power-Line Carrier System for Maximum Channel Utilization, **John A. Doremus.** *AIEE Transactions*, vol. 71, pt. III, Jan. 1952, pp. 429–40.

5. A Low Noise and Distortion Audio Multiplexing Equipment with High Stability Carrier Supply, **F. S. Beale.** *AIEE Transactions*, vol. 69, 1950, pp. 1028–38.

◆

Discussion

C. W. Boadway (The Hydro-Electric Power Commission of Ontario, Toronto, Ont., Canada). The technical presentation is an interesting engineering adventure. It would be of great interest and assistance to the communication field if, after completion of the over-all installation and field measurements as to line and equipment performance characteristics, the authors would prepare an additional presentation describing the practical difficulties and ultimate solutions involving the over-all application and installation.

The provision of communication and d-c signalling over 360 miles of virgin territory with numerous sidings, or tap points, is indeed a difficult problem. The text of this paper indicates that considerable thought and investigation were given to the problem before the choice of carrier on open-wire line was finally decided.

Because of the unusual length of the telephone line, and its exposure to a 23-kv single-phase power circuit, together with the numerous services and service taps required, it is expected that transmission and operating difficulties may be encountered before the system is finally accepted for service. There is a lack of power system information included in the technical presentation which prevents the reader from analyzing the inductive effects which may be expected from such a long power exposure with the telephone circuit.

In the section on "Initial Considerations" it was stated that the communication line was constructed with a transposition in the telephone circuit every fifth pole, or approximately 1,100 feet apart. It would be of great interest to know the magnitude of the induced voltage at various points along the line from the generating source, and whether or not additional remedial measures may be required for safe operation of the equipment and safety of operating personnel. In view of the fact that each siding has to use the physical line back to the dispatch center, it would be of interest if the authors were to elaborate on the magnitude of audio noise on the physical circuit, and the bridging loss at voice frequencies due to the numerous tap conditions; also, the effective transmission equivalent between the most distant siding and the dispatch center.

It is noted that the carrier equipment is to operate through a total line attenuation of approximately 50 decibels. We would assume from this estimated attenuation that it involves only the half-portion of line between repeater points, as our experience would indicate that the over-all installation would involve an average attenuation of 0.2 decibel per mile at approximately 100 kc not including equipment loss. On this basis, and including normal equipment losses (including line coupling apparatus, hybrids, and bridging filters), the over-all attenuation for the 360 miles would be of the order of 90 to 100 decibels. With this order of attenuation, it might be possible or necessary, due to noise over the carrier frequency spectrum, to consider additional repeater points.

The application deviates somewhat from our normal communication procedures as applied to telephone system facilities, and in view of the importance of the communication circuit requirements and the multiple services to be provided, I feel that the authors and those participating in the design and installation should be congratulated and encouraged in developing and applying a new communication technique.

A. K. Hansen, W. W. H. Dean, and **F. G. Mullins, Jr.:** Mr. Boadway suggests that further information be published when this installation is complete. This is contemplated, and will provide actual measured values of line noise, attenuation, and general operating characteristics. Starting in the fall of 1953, continuous noise and attenuation measurements will be made for a period of at least 1 year.

Calculations on the idealized line configuration give a very low interference figure in the communications line from the power line. Actual measurements with power on will be required to determine the effect of the unavoidable discontinuities introduced during construction. Drain circuits or other corrective steps can only be taken when the magnitude of the interference is known.

The effect of the many taps on the line is referred to. The coupling units at these taps are designed to introduce a bridging loss at carrier frequencies of less than 0.1 decibel and an insertion loss at voice frequencies of less than 2 decibels. In addition they present a low d-c series resistance to the railway signalling system.

The physical telephone circuit is presently in use over more than 100 miles of line with excellent performance. The effect of noise from the energized power line is yet to be determined, however.

Several calculated values are yet to be checked by actual measurement and undoubtedly unexpected problems will arise. We look forward to making a complete report on these when the project is completed.

Program Transmission Over Type-N Carrier Telephone

R. L. CASE
MEMBER AIEE

IDEN KERNEY
NONMEMBER AIEE

THE transmission of program material over a nation-wide system of wire-line networks interconnecting most of the country's broadcast and television stations has grown apace with the development of the radio and television arts. Today, program networks aggregating as much as 156,000 miles in length, including both radio broadcast circuits and television sound channels, may be employed during a single evening.

Initially, all program networks operated over voice frequency facilities, either open-wire or cable. The open-wire circuits were nonloaded, but the cable circuits, other than short runs in metropolitan areas used to connect studios and transmitters, were, in general, equipped with special loading, and the transmission variations with temperature were controlled by pilot-wire regulators. The voice frequency cable pairs assigned to program transmission were usually 16-gauge, so that the program levels in each repeater section might be maintained at adequate values, and the repeaters still spaced far enough apart to permit their being placed at the same locations as the message repeaters.

In 1945, program transmission on a commercial basis was first provided for each of the major networks over *K* carrier facilities between Los Angeles and Omaha, utilizing three message channels to provide a single program channel in each *K* system.[1] However, as only a few of the broadcasters requested facilities with a wider band than that provided by schedule *A* program (approximately 100–5,000 cycles) the program terminal was modified to permit one of the three channels to be returned to message service.

With the advent of the less expensive type-*N* carrier system[2] about 3 years ago, the possibility of both schedule *A* and schedule *C* (approximately 200–3,500 cycles) program transmission over one or two links of this facility became of great interest. This was partly because of the flexibility of the proposed arrangements and partly because of the economies envisioned. With the increased use of television facilities a high speed of propagation for the sound channel was required so as to synchronize properly with the video channel, a speed not possible with long voice frequency loaded pairs. In addition, in anticipation of the use of program over carrier, new cables, both quadded and nonquadded, were being fabricated without any 16-gauge pairs.

Schedule *A* and *B* (part time schedule *A*) program requires a band which is wider than the total frequency space allotted to a type-*N* channel. To obtain this band with existing type-*N* arrangements, it is necessary to give up three message channels. While schedule *C* and *D* (part-time schedule *C*) program requires the band space of only a single-message channel, it is necessary to widen the message band and to eliminate the 3,700-cycle built-in signaling system.

Engineering Considerations

Prototype program channel units providing schedules *A* and *C* transmissions were tried out on two Milwaukee-Madison experimental *N* systems in 1950. Exploratory work was also carried out to determine the feasibility of providing 8- and 15-kc facilities over these systems. As there were few commercial requirements at that time for program service with a band wider than that provided by schedule *A* facilities, it was decided to concentrate on the development of channel units for schedules *A* and *C* program transmission.

The Milwaukee field investigation indicated that while satisfactory program transmission was readily realized over type-*N* carrier systems in fair weather, excessive static interference resulted during severe electrical storms in the vicinity of the Milwaukee-Madison cable. Various remedial measures were tried on the *N* systems themselves, such as application of direct drainage on the *N* pairs; application of roof (low-pass) filters at the inputs of high-low repeaters and of cellar (high-pass) filters at the inputs of low-high repeaters, and substitution of a local battery supply for the individual repeaters in place of the power supply usually fed over the *N* pairs. These measures offered only minor reductions in the static disturbance, so it was concluded that most of the noise was introduced through the open-wire taps along the cable. However, because of the large number of taps involved it was not found practicable to apply suppression measures to them during the 1950 static season.

The following year, another series of tests was conducted, this time on an *N* system operating over the Harrisburg-Sunbury, Pa., toll cable, which, when the field study began, had only a single open-

---

Paper 53-335, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Pacific General Meeting, Vancouver, B. C., Canada, September 1–4, 1953. Manuscript submitted May 25, 1953; made available for printing July 13, 1953.

R. L. CASE and IDEN KERNEY are with the Bell Telephone Laboratories, Inc., New York, N. Y.

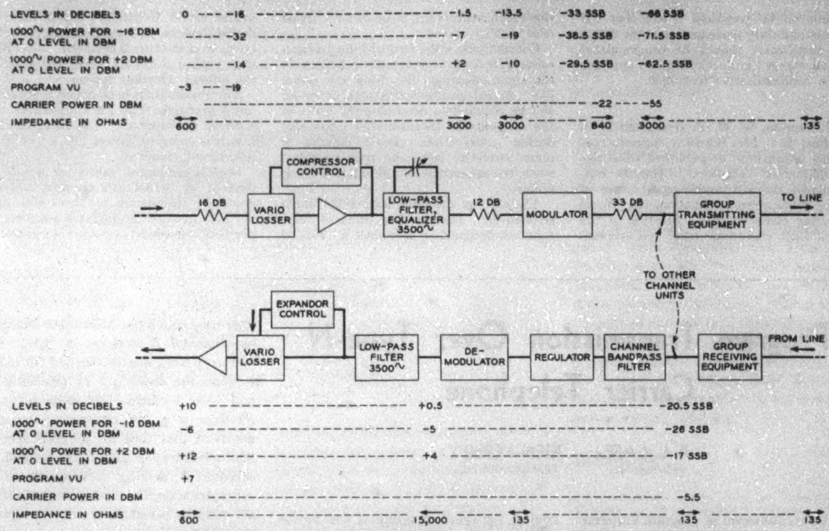

LEVELS IN DECIBELS
0 --- -16 --- -1.5 -13.5 -33 SSB -66 SSB
1000~ POWER FOR -16 DBM AT 0 LEVEL IN DBM
-32 --- -7 -19 -38.5 SSB -71.5 SSB
1000~ POWER FOR +2 DBM AT 0 LEVEL IN DBM
-14 --- +2 -10 -29.5 SSB -62.5 SSB
PROGRAM VU
-3 --- -19
CARRIER POWER IN DBM
-22 --- -55
IMPEDANCE IN OHMS
600 3000 3000 840 3000 135
COMPRESSOR CONTROL
16 DB
VARIO LOSSER
LOW-PASS FILTER, EQUALIZER 3500~
12 DB
MODULATOR
33 DB
GROUP TRANSMITTING EQUIPMENT
TO LINE
TO OTHER CHANNEL UNITS
EXPANDOR CONTROL
VARIO LOSSER
LOW-PASS FILTER 3500~
DE-MODULATOR
REGULATOR
CHANNEL BANDPASS FILTER
GROUP RECEIVING EQUIPMENT
FROM LINE
LEVELS IN DECIBELS
+10 --- +0.5 --- -20.5 SSB
1000~ POWER FOR -16 DBM AT 0 LEVEL IN DBM
-6 --- -5 --- -26 SSB
1000~ POWER FOR +2 DBM AT 0 LEVEL IN DBM
+12 --- +4 --- -17 SSB
PROGRAM VU
+7
CARRIER POWER IN DBM
-5.5
IMPEDANCE IN OHMS
600 15,000 135 135 135

Fig. 1. Block schematic of schedule C compressor and expandor, level diagrams included

wire tap location with two open-wire pairs entering the cable at that location. Means were provided at this junction for switching suppression filters in or out of the circuit as desired, one filter being supplied for each of the two open-wire pairs.

The observed effectiveness of the suppression filters located at this tap was found to be about 20 decibels (db), a value great enough to permit satisfactory schedule *C* program transmission even under storm conditions. While no tests have been made on schedule *A* performance when a severe electrical storm was centered over one or more of the open-wire taps, it is expected that such suppression will also permit the use of these wider band program facilities during periods of high static noise. To obtain satisfactory year-round program service over *N* systems, it may be necessary to treat each and every open-wire pair at its junction with the cable carrying the *N* pairs.

Type-*N* carrier will provide more economical program circuits than can be obtained with voice frequency facilities, which, though meeting the technical requirements, are inflexible. For example, in planning for schedule *A* program service over such voice facilities, a sufficient number of program pairs must initially be fabricated in the cable, not only to take care of immediate requirements but also to provide for future growth during the life of the cable. In addition, the program pairs must be specially loaded (*B-22* or *B-44*), and, if the toll cable in which these pairs are located is aerial, pilot-wire regulation must be furnished for the associated program amplifiers along the circuit, as well as for the 2-wire and 4-wire message repeaters. This can be an expensive addition when only a small number of voice circuits is involved.

The design of the *N*-carrier-derived schedule *A* program system makes use of all of channel 6 and of a 1-kc band in channel 5 and in channel 7. If there is a paralleling *N* system operating in the same cable, crosstalk which might result from the 3,700-cycle signaling tones of that system, could show up at the receiving schedule *A* terminal as a 3,700-cycle tone effected by the channel 6 upper and lower side bands of 3,700 cycles; a 4,300-cycle tone effected by the channel 5 upper side band, and a 4,300-cycle tone effected by the channel 7 lower side band. The disturbing effect of these tones is of a different nature when several *N* systems are operating in the same cable, in that the 3,700-cycle signaling oscillators of the various *N* systems are not in synchronism and are random in frequency at about the 3,700-cycle point. To add to this randomness, the channel carrier frequencies and the repeater carrier frequencies in the various systems in the cable also differ from each other. If there is also an *ON* system[3] operating in the same cable, crosstalk from one of its carriers could be transmitted through the program terminal and could show up as a 4-kc tone. When more than one *ON* system is operating in the cable, the effect of the 4-kc tones may be somewhat like that of the interference from multiple *N* systems just described. The results of a recent field study of schedule *A* program transmission operating over *N* carrier in a nonquadded cable between Bangor and Presque Isle, Me., showed that there was a tendency for the disturbance from the 3,700-cycle signaling tones of other *N* systems in the same cable to approach resistance noise in character. The values of the total interference from these random tones were found to be acceptable even over such a long nonquadded cable. While there may be marginal cases of *ON* interference,

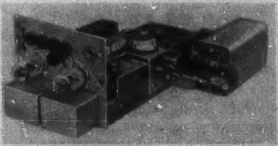


Fig. 2. Schedule C compressor subassembly

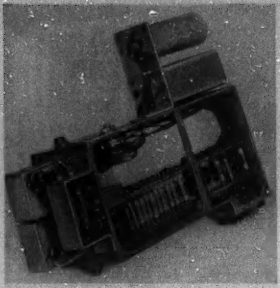


Fig. 3 (left). Schedule C expandor subassembly

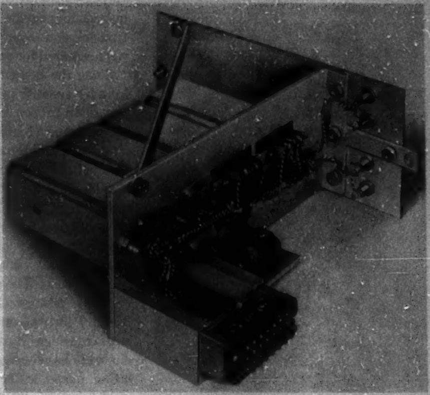


Fig. 5 (right). Schedule A low-frequency subassembly

it is believed that, in general, satisfactory crosstalk values may be realized by the selection of pairs or by a reduction in *ON* carrier level, or both, when a nonquadded cable is employed. The use of quadded toll grade cable should permit the operation of other *ON* or *N* systems in the same cable as schedule *A* program, without any special precautions, and with a reasonable assurance of low levels of disturbance present in the program channels.

Design of Program Channel Units—General

By making certain modifications in the type-*N* terminal channel unit, it is possible to replace any type-*N* message channel unit by a modified unit, and provide schedule *C* program service over that particular channel. One or more schedule *C* program channels may be used on a type *N* system.

To transmit the higher grade schedule *A* program over an *N* system, however, requires more elaborate changes. In the present design three specific message channels of the *N* system are relinquished and are replaced by one schedule *A* program channel. However, one message channel may be recovered by using the space of one of the channel units for the addition of a higher frequency channel unit now under development. Provision is made for the application of only one schedule *A* program circuit to an *N* system. External equalizers, predistorting and restoring networks, and compandor with a wider frequency band are required. All these items are presently available. No modifications of any kind are necessary in the type *N* group units or repeaters for either schedule *C* or *A* program.

Schedule *C* or *D* Program Channel Unit

Like the type-*N* message channel units, the schedule *C* program channel unit consists of a compressor subassembly, an expandor subassembly, and carrier subassemblies. The program compressor and expandor subassemblies are the same for any one of the 12-channel units. The carrier subassembly with its carrier frequency modulating oscillator and band filter is different for each channel unit but is the same for message and program units.

The compressor and expandor subassemblies, however, were considerably modified for program use. Since the program circuits are always operated on a 1-way basis, the 2- or 4-wire terminating equipment was removed and the transmitting low-pass filter and equalizer on the compressor subassembly were modified to extend the band down to 200 cycles and up to 3,500 cycles. An adjustable feature is provided in the high-frequency part of the equalizer to take care of variations that may be present in the over-all circuit. All signaling equipment was removed from the expandor subassembly and the receiving low-pass filter was modified to extend its band to 3,500 cycles.

Additional filtering was provided in the plate circuit of the compressor and expandor so that different programs could be transmitted in both directions simultaneously over the same channel unit with negligible crosstalk between them.

A block schematic of the compressor and expandor subassemblies is shown in Fig. 1, and views of the equipment in Figs. 2 and 3. Minor alterations were made in the message unit die castings, and simple brackets were added to accommodate the additional circuit elements.

A typical over-all transmission frequency characteristic for a schedule *C* type-*N* link is shown in Fig. 4. It will be noted that the characteristic is flat within ±1.5 db over the 200- to 3,500-cycle band.

When the internal noise generator provided in the *N* system to mask crosstalk is adjusted to give +33 decibels adjusted of noise on message circuits at the +10 level point, the noise on the schedule *C* program channel at the +8 volume-unit point is well under a reasonable objective for this class of service. The crosstalk between message and program channels or from program-to-program channels is negligible.

The program unit mounts interchangeably with any message unit. To provide program service over a particular channel, it is only necessary to remove the message channel unit from its plug-in

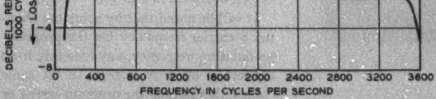


Fig. 4. Typical transmission frequency characteristic of schedule C type-N single link with ten intermediate repeaters

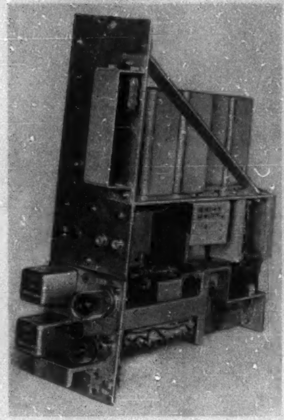


Fig. 6. Schedule A channel unit consisting of low-frequency and carrier subassemblies

socket and replace it by the program unit having the same channel number. If desired, the carrier subassembly may be removed from the message unit and reassembled with the program compressor and expandor subassemblies to form the new program channel unit.

The heater power required for the program channel unit is less than for the message unit but the program unit has a resistance load added so that the total heater battery drain for the program unit is the same as for the message unit. Thus no readjustment of the resistances in the heater battery supply circuit is necessary when changing from message to program units or vice versa.

Schedule *A* or *B* Program Unit

The transmission characteristics of the compressor and expandor of the type-*N* channel unit make them unsuitable for the wider band of the schedule *A* program. Likewise, the 4-kc wide-channel band-pass receiving filter is unsatisfactory for 5-kc transmission. Consequently, more elaborate changes are required to provide schedule *A* program transmission over type-*N* carrier than were required for schedule *C*.

Since the demand for schedule *A* service over type-*N*, initially at least, is expected to be considerably less than that for schedule *C*, arrangements have been made to provide only one 5-kc program channel over a type *N* system. This is obtained by removing message channel units 5, 6, and 7, and by replacing these three units with one program unit having a carrier subassembly similar to channel 6 but with a 5-kc wide-band receiving filter. The program band now occupies a 1-kc portion of both channels 5 and 7 but of course does not interfere with message transmission since those channels have been removed. For regulating purposes the transmitted carrier of channel 6 is increased 5 db in order to maintain the same total transmitted carrier power of the 12 channels as existed when channels 5, 6, and 7 were used for message service. With the channel 6 carrier increased 5 db, the program input to the modulator is increased a corresponding 5 db in order to gain signal-to-noise improvement.

The message compressor and expandor subassemblies are replaced by a single unit consisting of an equalizer and transformers so that the 600-ohm sending and receiving voice frequency program circuits may be properly coupled to the modulator and demodulator circuits on the carrier subassembly. These units are shown in Figs. 5 and 6.

The channel 6 unit is built out to take the same heater load as the corresponding message unit. To avoid the necessity for any readjustment of heater voltage, plates are provided, each of which is equipped with a resistance to take the same heater load as would be taken by a message channel unit 5 or 7. These plates may be plugged into the blank spaces caused by removing channel units 5 and 7. Fig. 7 shows a plate with heater load resistance.

The schedule *A* program unit can of course be readily plugged into the place of the channel 6 message unit. A view of a type-*N* terminal equipped with a schedule *A* channel unit is shown in Fig. 8. The engineer in the photograph is inspecting a schedule *C* channel unit.

Fig. 8. Type-N terminal equipped with schedule A channel unit and plates. Engineer inspecting schedule C unit

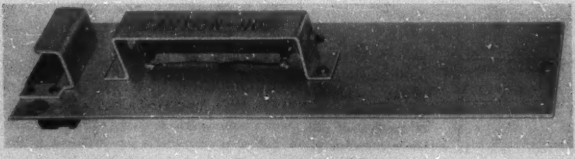


Fig. 7. Plate with heater load resistance

The 1*A* compandor is used external to the *N* system to provide part of the improvement in noise and crosstalk performance desired for schedule *A* program. The 1*A* compandor has a range of compression and expansion similar to the type *N* compandor but has a bandwidth suitable for the schedule *A* program.

Further improvement in noise and crosstalk is obtained by the use of the 200-type predistorter and restorer, devices to partially equalize the power of the low and high frequencies at the sending end, and restore them to their normal relationship at the receiving end. A 5-kc low-pass filter is placed at the transmitting end to prevent extraneous frequencies above the program band from causing interference into the adjacent type *N* channels. Equalizers are furnished at the receiving end as required to meet the over-all transmission objective.

A block schematic of the over-all schedule *A* program system, together with a level diagram, is shown in Fig. 9. The circuit is arranged so that different programs can be transmitted simultaneously in the two directions. With the arrangement shown in Fig. 9 it is expected that the transmission will be flat within ±1 db over the band from 100 to 5,000 cycles, as shown in Fig. 10.

It will be noted that by using the channel 6 carrier frequency for the program-modulating frequency, any crosstalk from carriers of other *N* systems in the same cable either falls on the program carrier or

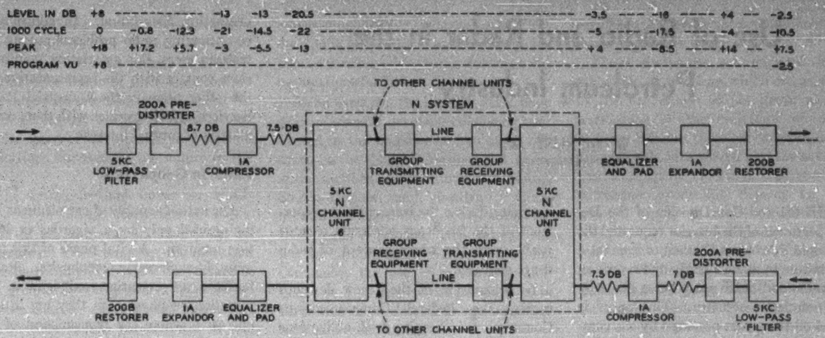
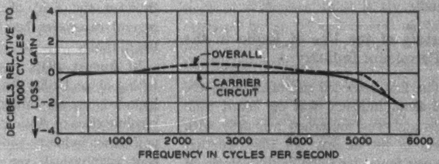

LEVEL IN DB +8 -13 -13 -20.5 -3.5 -16 +4 -2.5
1000 CYCLE 0 -0.8 -12.3 -21 -14.5 -22 -5 -17.5 -4 -10.5
PEAK +18 +17.2 +5.7 -3 -5.5 -13 +4 -8.5 +14 +7.5
PROGRAM VU +8 -2.5
TO OTHER CHANNEL UNITS
N SYSTEM
200A PRE-DISTORTER
8.7 DB
7.5 DB
LINE
5KC LOW-PASS FILTER
1A COMPRESSOR
GROUP TRANSMITTING EQUIPMENT
GROUP RECEIVING EQUIPMENT
EQUALIZER AND PAD
1A EXPANDOR
200B RESTORER
5KC N CHANNEL UNIT 6
5KC N CHANNEL UNIT 6
GROUP RECEIVING EQUIPMENT
GROUP TRANSMITTING EQUIPMENT
200A PRE-DISTORTER
7.5 DB
7 DB
LINE
200B RESTORER
1A EXPANDOR
EQUALIZER AND PAD
1A COMPRESSOR
5KC LOW-PASS FILTER
TO OTHER CHANNEL UNITS

Fig. 9 (above). Block schematic of schedule A program system

Fig. 10 (right). Typical transmission frequency characteristic of schedule A type-N single link. Over-all characteristic includes external equalizers, predistorting and restoring networks, compandor, and 5-kc low-pass filter



falls outside the program band. A rearrangement of the carrier modulator of the schedule *A* program system, so as to employ only two channels of band space instead of three in the *N* system, would have introduced more co-ordination problems with the carriers of the other *N* systems.

Reversal Arrangements for Schedule *A* Program

Although program over *N* carrier is usually limited to the last one or two links of a program network, provision for reversing must be considered, since all or part of the particular program may originate at the end of the network, for example, when the President of the United States on tour broadcasts from his train while it is held on a siding at some small locality. Circuit arrangements to provide for this reversing are now under development. With these arrangements, only one direction of transmission of the program at a time is used, the direction at any time being controlled by the customer or telephone company as required.

The relay-reversing arrangements for the voice frequency part of the circuit are similar to those used for other standard program circuits. To reverse the carrier part of the program circuit, a new unit is being designed to be used in place of channel 5. This unit will have a channel 5 carrier oscillator of somewhat lower amplitude than the normal message carrier so that the transmitted tone will not appreciably affect circuit regulation. This new channel 5 carrier oscillator will provide the transmitted reversal tone under control of the transmitting directional setup. Transmission of the channel 5 carrier oscillator tone over the circuit conditions the circuit for transmission in that direction. A crystal pick-off filter and receiving circuit are used to identify the tone sent by the distant end. The receiving channel regulator is eliminated to avoid large increases in receiving again with possible false indications when the reversing tone is removed at the sending end. The channel 5 reversing unit is built out to take the same heater load as the corresponding message unit and can readily be plugged into the channel 5 space.

Typical Bell System Application

A number of schedule *C* and *D* program systems are already in operation. It is expected that by the end of 1953 schedule *A* and *B* program transmission will be operating over *N* carrier systems between Joplin and Springfield, Mo.; Santa Ana and Los Angeles, Calif.; and Duluth and Minneapolis, Minn.

References

1. A System for 8,000-Cycle Program Transmission. *Electrical Engineering*, vol. 68, May 1949, pp. 395–400.

2. A New Telephone Carrier System for Medium-Haul Circuits, R. S. Caruthers, H. R. Huntley, W. E. Kahl, Ludwig Pedersen. *Electrical Engineering*, vol. 70, Aug. 1951, pp. 692–97.

3. Type-*ON* Carrier Telephone, R. D. Fracassi, H. Kahl. *AIEE Transactions*, vol. 72, pt. I, 1953 (Jan. 1954 section), pp. 713–21.

No Discussion

Use of Radio and Radar in the Petroleum Industry

W. M. RUST, JR.
NONMEMBER AIEE

THE Oil and Gas Division of the Department of the Interior requested the National Petroleum Council to appoint a committee to make a study and report on the radio and radar usage being made by the petroleum industry. This paper is based on the report prepared by the committee, under the chairmanship of F. W. Littell.

History of the Use of Radio in the Petroleum Industry

The first recorded use of radio in the petroleum industry was on board tank ships and was concerned primarily with safety of life at sea. In 1925, another use was found for radio in the petroleum industry by geophysical crews engaged in prospecting for petroleum deposits.

In 1928, needs for radio communication in the industry led to the submission of applications to the Federal Radio Commission (the predecessor to the Federal Communications Commission) for radio licenses by several oil companies. The Commission ruled at that time that radio channels could not be assigned for private use, but suggested that it would be in order for the companies interested to form a public service corporation to accomplish the desired result. Accordingly, in January 1929, the Petroleum Communications Corporation, later renamed the Western Radio Telegraph Company, was formed.

In 1932, a few oil companies began to use radio in the Special Emergency Radio Service on a very limited basis in the swamps of Louisiana for communications between drilling rigs and base offices. In 1938, a new class of station, "provisional," was established by the Federal Communications Commission under their rules governing "Miscellaneous Radio Services." A number of petroleum companies qualified for this service, because its primary use was for operations involving the safety of life and property. By 1941, considerable interest was being evidenced in the possibilities of radio usage throughout the industry, but World War II prevented any expansion of radio installations.

Because of the technical advances made in the radio art, under pressure of military necessity during the war, entire new portions of the frequency spectrum became usable, and the whole concept of radio usage changed.

In recognition of these new developments, the Federal Communications Commission, in keeping with obligations to study new uses of radio, decided to broaden the field of users to include several categories heretofore denied the use of radio.

In August 1946, the Commission established the Utility Radio Service, which included, among others, the petroleum-pipeline class of station. However, other petroleum industry radio operations, except geophysical, were still classed as "experimental."

Realizing the importance of radio as an operating tool for all branches of the industry, the Petroleum Industry Electrical Association petitioned the Commission in 1947 for the establishment of a "New Radio Service Governing Radio Stations Used in Petroleum Industry Production and Pipeline Operations and Promulgation of Rules and Regulations Therefor." Steps were taken to interest the American Petroleum Institute in the formation of a radio committee which could encompass all phases of the industry's varied operations in a recognized association which was truly representative of the industry as a whole. The newly created Central Committee on Radio Facilities of the American Petroleum Institute immediately concerned itself with the most pressing problem, that of finalizing the establishment of the Petroleum Radio Service.

Finally, on May 3, 1949, the Commission announced the adoption of its Industrial Radio Service Rules in final form, to become effective July 1, 1949. Subpart *G* of part 11, "Rules Governing Industrial Radio Services," covered "Petroleum Radio Service" and included all of the major target goals contained in the original petition for the service.

Upon establishment of this new service, another problem, that of co-ordination of frequencies assigned to the service, became of paramount importance. The result was the formation of the National Petroleum Radio Frequency Co-ordinating Association. In addition to co-ordinating radio problems among petroleum users, this association maintains close contact with the representatives of the other users in the Industrial Radio Services and co-operates with them on all matters of mutual interest.

Radio in Geophysics

The nation's supply of petroleum can be maintained only by a vigorous exploration program. A vital phase of this program is exploration geophysics. Radio facilities are essential auxiliaries in exploration geophysics, as they are in any operation where the transmission of information of any kind is of vital importance. While, in general, they are but one link in the rather diverse instrumentation that makes up a geophysical exploration unit, in certain cases they are a primary link.

Just as oil or gas might be collected in an inverted bowl held over water, so the oil and gas content of a permeable rock stratum will gravitationally segregate and be trapped at the top of a closed fold in the stratum, provided that the overlying rock is effectively impermeable. The location and delineation of such features is the ultimate goal of geophysical exploration.

It is self-evident that any geophysical technique must be based on distinguishable variations in the physical characteristics or properties of the subsurface rocks. They vary in:

1. Density;
2. Magnetic properties;
3. Elastic characteristics;
4. Electrical properties.

Electrical methods are of considerable importance in mineral exploration, but have very limited application to date in petroleum exploration.

The operational procedure based on density variations comprises nothing more than the transportation of a suitable gravity meter from point to point over the area to be investigated. As in the case of gravity measurements, the field procedure based on variations in magnetic properties comprises nothing more than the taking of precise magnetic measurements from point to point over the area under investigation. Airborne magnetometer operations are unique in geophysical oper-

---

Paper 53-381, recommended by the AIEE Radio Communications Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Kansas City, Mo., November 2–6, 1953. Manuscript submitted June 22, 1953; made available for printing July 29, 1953.

W. M. Rust, Jr., is with the Humble Oil and Refining Company, Houston, Texas.

ations in that the necessary measurements are made in transit. Herein, of course, lies the particular advantage of this reconnaissance method, since the required magnetic data are collected at a speed determined only by the practical limitations on the speed of the aircraft, and with no limitations imposed by the nature of the terrain.

The seismic method makes use of differences in elastic properties of rocks. First of all, the seismic wave must be generated. This comprises the initiation of a shockwave at or near the surface of the ground by the use of a suitable explosive charge. An essential element in the method is the transmission of the shot instant so that it may be recorded on the same record as the seismic events.

The recording truck carries all of the instruments required for the recording of the seismic events picked up by suitable geophones or seismic detectors comprising the receiving array. Supplementing this equipment are, of course, the necessary communication facilities.

Communication equipment is necessary, especially for the transmission of the shot moment, since the useful information on any seismic record is the time interval between the shot and the arrival of reflected or refracted events. The desired timing accuracy is on the order of 0.001 second; hence, good recording reliability is essential. Voice communication is also necessary to co-ordinate operations between recorder and shooter which may involve distances of a few thousand feet up to ten or twelve miles, depending upon the seismic technique in use.

Radio facilities are required for geophysical operations primarily for three purposes. The first is the co-ordination of activities, including safety measures. The second is the transmission of special data. The third is the determination of position by radiolocation methods. The extent to which a specific geophysical crew utilizes radio varies widely with the nature of the geophysical technique employed and with the type of terrain covered. Radiolocation is the newest, but by no means the least important. It is fair to say that airborne and offshore operations are practical only when radiolocation techniques are employed.

Use of Radio in Production Operations

Petroleum is useful only when it has been drilled and produced. It has long been recognized that efficient communications are essential in drilling and production operations, because of the inherent hazards of such operations. Production is frequently found in remote regions, far removed from any existing communications facilities.

As an example, the last few years have seen substantial developments being conducted in the marshes of Louisiana, offshore in the Gulf of Mexico, and in remote sections of Montana, Wyoming, South Dakota, and elsewhere. Communication with drilling locations in these areas is possible only through the use of automobiles, boats, and radio. Obviously, the use of automobiles and boats as a means of communication in these areas is so slow as to be completely useless in emergencies.

In an emergency, it may take many hours for the toolpusher to drive to the nearest telephone and get a call through to his central office or to a point where help is available. During these lost hours untold damage may be done to the well and the drilling equipment on the well. If the rig is equipped with radio, immediate contact can be established and the loss prevented or lessened. The same conditions exist where drilling is being conducted offshore. In these offshore operations, the personnel on the rig can be kept advised at all times about weather conditions, and, in the event of severe storms or hurricanes, warnings can be sent to the drilling site in time to allow for proper precautions to be taken before the blow strikes. This is possible only where radio communication is available.

Radio contact with drilling rigs has also proved valuable from another angle. Very often data from the well make it essential to change the drilling procedure and, in many cases, because of communication difficulties, the instructions do not reach the rig until after it is too late to make the change. No doubt more than one well has been ruined because of the failure to get proper instructions to the rig in time. The use of radio greatly reduces this hazard.

Based on the actual experiences of many of the companies now using radio, it has been conclusively demonstrated that radio is the only practical solution to many of the communication problems of producing operations.

In addition to those companies actually engaged in drilling and production operations, there are many hundreds of companies engaged in supplying these companies with vital materials and services. For the most part, these companies serve the drilling companies and oil companies exclusively, and they are, therefore, considered part of the petroleum industry, even though they do not actually hold the leases or produce the oil.

These companies are engaged in a multitude of activities which are as vital to the production of petroleum as the work done by the individual oil companies. Because these service companies are so closely interwoven into the over-all petroleum operation, and because so many petroleum operations depend on the prompt supply of needed materials or services, they began using radio almost as soon as the production and drilling companies. Without minimizing the value of radio the drilling or production company, the service company often has an even greater need for radio communication.

Indicative of the size of the radio operations of such companies is the system operated by an oil-well cementing company. This 2-way radio system is one of the largest privately owned 2-way communication systems in the world. This system, consisting of dozens of base stations and several hundred mobile units, stretches from near the Canadian border to the Mexican border.

Pipeline Use of Radio

The pipeline industry is not a new one; therefore, the communications problems of the pipeline industry are not new, though they have changed greatly since the first short low-pressure crude-oil pipelines were built. Around the turn of the century, when pipelines were still new but useful adjuncts to the young oil industry, private communication systems for pipelines were necessary simply because public communications were not available to the remotely located pump stations. As the oil industry grew, and with the discovery of oil in Oklahoma, California, Louisiana, and Texas, pipelines reached out into even less densely settled areas, and in many of these, the communication facilities erected by the pipelines were the only facilities available to anyone. Generally, these conditions do not exist today. However, it is still the usual practice of the pipeline industry to own and operate its own communications facilities, since even today many of the pump stations are remote and because reliable full-time communication among all points is an operating necessity.

To understand the communications problems of pipelines, it is necessary to realize the magnitude of the shipping problem handled by a large line. A 250,000-barrel-per-day line 1,000 miles long moves approximately 37,500 tons per day through the line, for a total movement of product of 37,500,000 ton-miles per day. This is equivalent to a railroad keeping eighty 60-car trains in constant service.

On a tight line, with no intervening tankage anywhere on the line, any interruption to the flow at any point will be felt at every other point. If the flow is not reduced, the pressure in the line will continue to rise until the line is split, or pump units start breaking down. From this it can be seen that a tight line must have available a communication system as reliable as can be had and flexible enough to afford instant communication among the pump station engineers at all times. In addition to the constant interchange of operational information among pump stations, the dispatcher must be kept informed as to the passing of the head end and tail end of the various batches of oil in order that he can keep his dispatching orders corrected.

A communication requirement peculiar to pipelining is the keeping of the "line balance." The quantity being pumped into the line by the first station should equal the amount being delivered by the line at its terminus. When the balance shows "short" for two or three checks in succession, it is reasonable to suspect that the oil is being lost through leakage. Considering that a 22-inch line can deliver up to 10,000 gallons of crude oil per minute through a split, the essentiality of good communications is seen immediately.

Equally as important as the oil pipelines are the natural gas pipelines. Since gas is a fuel that is not readily adaptable to conventional storage methods, controlled stream flow must be utilized to accommodate varying demand. Pipeline pressure as well as pumping rate must be skillfully controlled to insure deliverability consistent with public requirements. The most important factor in controlled flow is a reliable communication system. Skilled dispatchers must watch demand along the system, varying gas intake from numerous well heads to accommodate these fluctuations, and in many instances, interrupting supply to less critical users in order to prevent hardships developing in areas where rapid temperature drops suddenly increase demands. Rapid order changes are necessary from the dispatching experts to handle these situations.

In some pipeline systems, the dispatching expert receives the pertinent telemetered information as to pressures, flow, and input quantities into the pipeline, delivered quantities from the pipeline to consumers, and other data, via microwave into headquarters. With this system of operation, the dispatcher continuously observes line conditions and, when system changes are required, either uses remote-control functions over the microwave facilities to effect the desired changes, or else utilizes voice channels to issue the required orders.

The far flung activities of a gas or oil pipeline result in very-high frequency (VHF) communications problems that can not be handled satisfactorily without supplementing the relatively short range of the VHF equipment with other auxiliary communications facilities without such range restrictions. It is not unusual to require that maintenance crews a hundred miles or so apart be co-ordinated for maximum efficiency and safety.

The petroleum industry has adopted microwave, one of the fastest growing communications mediums ever to be developed in this country, as a solution to this and other long-range communication problems. Many such systems are in operation and more are being installed.

Utilization of those frequency bands which the Federal Communications Commission has made available for microwave systems on a developmental basis has been great in the petroleum service and, in particular, in the long-haul natural-gas pipelines. At present, some 22 companies have approximately 15,000 miles in operation. Since these systems have from 8 to 24 circuits each, the total circuit miles are many times this figure.

In addition to the communications flexibility offered by microwaves, the simple addition of the microwave system's multichannel capabilities to the tools available to the pipeline industry is destined to change the industry more than anything since the introduction of the electric centrifugal pumping unit. Once a microwave system is installed, there are a very large number of signaling channels available for supervisory control and remote indicating applications which do not substantially affect the ability of the system to carry communications signals. This added capability of microwave systems, which is available at a very low cost after the system has been installed, is such as to make the completely remote operation of pump stations economically feasible for the first time.

Shipboard Use of Radio

While pipelines constitute the backbone of the petroleum transportation system, ships and barges also perform a vital service in moving petroleum economically. Marine operators moving petroleum products have been very progressive in taking advantage of the aids offered through radar and radio communication. The character of their cargo has made such help especially useful to them.

The petroleum tankers on the high seas use radio and radar in the same manner as other vessels, and are distinguished only by their greater utilization of these aids. They have proved to be an invaluable part of the tankers' equipment.

If possible, radar and radio are even more important to the operator on inland waters. The importance of the radiotelephone to the inland waterways transportation industry may be visualized by reference to an example of its practical use. The personnel of inland waterway vessels usually make soundings of all dangerous crossings and locations on the rivers, and these findings can be relayed to other vessels approaching such channels. Likewise, channel information furnished by the United States Corps of Engineers and the United States Coast Guard can be passed on by radiotelephone to vessels immediately upon its becoming available. This is a highly essential function when it is realized that river channels do not remain static because of the continuous shifting of sandbars and consequent channel changes. The importance of radar in this service is indicated by the existence of more than 300 radar-equipped vessels on inland waterways.

Petroleum Radio Relay Stations

The petroleum industry has many applications of radio where communications are needed over distances greater than can be covered directly from one fixed or base station to another fixed, base, or mobile station. In these cases, relay stations must be used to extend the range to accomplish the desired results. Very often mobile stations must communicate with each other, but they can not do so directly because of the terrain or the distance involved.

A pipeline approximately 185 miles long through rough mountainous country has been operated for several years with a VHF radio-relay system. This pipeline goes through territory where it would be practically impossible to maintain a wire-telephone line during the heavy snow season. The radio-relay system provides adequate communication with the mobile units not only along the pipeline but also traveling along the roads remote from the pipeline. Repair trucks and snow cats can keep in touch by means of radio when they are out in hazardous weather conditions, and they can call for material or help of any kind when needed. Maintenance foremen can reach pump stations promptly to restore the line to service after repairs have necessitated its being shut down. This greatly assists in maintaining the flow of oil essential to the industrial and governmental activities served by the pipeline. No other system of communication would have served this pipeline as well, and without it the hazards to life and property would be far greater than now exist.

Refinery Use of Radio

The use of radio communication within the refineries of the oil industry has grown with phenomenal speed in the last few years. At first thought, one might conclude that even the larger refineries would find point-to-point wire communications adequate for their needs. An average sized refinery of 100,000-barrel-per-day capacity will encompass an area of 1,000 acres, and be manned by 1,500 to 2,000 people. To carry out their duties most of these people must be in motion between unspecified locations. This presents a communication problem that is unique.

The largest section of manpower that benefits by mobile radio is that involved in maintenance and repair. The maintenance of plant equipment in safe and operable condition is the job of over half the manpower in a refinery. Every truck, every car, essentially every vehicle, is in constant 2-way communication with a central dispatcher.

A typical network for maintenance and construction in an average refinery with 100,000-barrel-per-day processing capacity would consist of a base station and approximately 100 mobile units. These mobile units would be located in supervisors' cars, craft foremen's pickup trucks, transportation department trucks, and straddle carriers. Actual statistics from existing systems show that on such a system the total number of messages handled in an 8-hour day will be between 800 and 900. During peak operating periods, the call density will reach one call every 30 seconds.

Regardless of how well planned a refinery may be, regardless of the intricate measures taken to avoid disaster, there comes the day when a serious emergency occurs. Point-to-point communications on what well may be a rapidly changing field are needed with instant dispatch. To meet such emergencies, the users of refinery radio have worked out systems and procedures in which the radio network is an essential part.

Use of Radio in Marketing Operations

Radio is being put to extensive and rapidly increasing use in the marketing and distribution of petroleum products. This use of radio in marketing operations has developed almost entirely since the establishment of the Highway Truck Radio Service on July 1, 1949. Prior to that date, radio for highway trucks was licensed on an experimental basis only and was in extremely limited use.

One of the largest users of radio in the Highway Truck Radio Service is the Liquified Petroleum Gas Communications Association. This association is comprised of 38 liquified petroleum fuel dealers in the state of Arkansas who have banded together for the sole purpose of providing a radio-communication service to the members. The association holds authorizations from the Federal Communications Commission for 38 base stations and 200 mobile units.

Petroleum Use of Radar and Other Radiolocation Methods

The petroleum industry adapted radar and shoran, which were developed by the military forces during World War II, to their own use immediately at the close of the war. They have provided modifications of both these two systems and have introduced a third system, the phase-comparison system, that is peculiarly suited to their specific needs.

The petroleum industry has made four diverse uses of radar in its operations. These are:

1. As a navigation device and as an anti-collision device on board the tankers of the petroleum industry's fleet;
2. As a location method for determining the position of the geophysical survey boats on the Continental Shelf off the coasts of the United States;
3. As a harbor surveillance device for guiding the boats for marine drilling operations into and out of coastal harbors during periods of low visibility resulting from fog;
4. As a device for detecting changes in weather, such as line squalls or hurricanes.

The second type of radiolocation device used in the petroleum industry has been Shoran. Shoran was developed for the guidance of aircraft engaged in precision bombing and is particularly applicable to aeronautical use. It has been used in the petroleum industry in connection with geophysical exploration, either the airborne magnetometer or the waterborne operations on the Continental Shelf. In many of the geophysical operations on the Continental Shelf, distances of 100 miles or more are involved. This precludes the possibility of using pulse type radiolocation methods for such surveying purposes, except for airborne operations, since all practical pulse systems operate in the frequency bands where propagation is restricted to line-of-sight distances. To overcome this limitation, the petroleum industry has employed a third type of radiolocation method, the phase comparison method. This method uses a continuous wave signal at a medium frequency which can be transmitted far beyond the horizon.

National Petroleum Radio Frequency Coordinating Association

The National Petroleum Radio Frequency Coordinating Association was organized in Washington, D. C., May 16, 1949. A tabulation of radio stations eligible for operation in the Petroleum Radio Service as of June, 1949, listed approximately 5,000 stations.

During the year June, 1949, to June, 1950, approximately 360 users requested frequency recommendations covering 10,450 transmitters. These were for all phases of petroleum operations, including the rapidly expanding products and natural-gas pipeline systems.

During the year 1950 to 1951, the number of users increased to a total of 438 and the number of recommendations to 15,603. This growth continued so that by June 30, 1952, there were 553 users with recommendations for 21,400 transmitters.

Forty-seven per cent of the petroleum radio stations are concentrated in Texas, Louisiana, Arkansas, and Mississippi. A belt approximately 75 miles deep along the Texas and Louisiana Gulf Coast contains the major portion of this concentration. The Petroleum Radio Service has reached the point of saturation in many sections of this area, with as many as 16 users operating co-channel, and with four of these within the same immediate area. This situation is not confined to any one band or group of frequencies. Technical improvements in equipment design have made practical adjacent channel installations within the same area; otherwise the limited number of frequencies in the Petroleum Radio Service would not have accommodated the large number of users.

At the annual meeting of the National Petroleum Radio Frequency Coordinating Association in Washington, D. C., June 7, 1951, a special committee on 20-kc equipment was appointed to conduct tests to determine whether the state of development of mobile radio communication equipment was such that operation on adjacent channels, with 20 kc center separation, is feasible in the bands from 25 to 50 megacycles.

It was concluded from these tests that

with proper adjustment of the center frequency and maximum deviation of the transmitters, it will be possible to operate adjacent channels with 20-kc separation with commercially available equipment, with practical geographical spacings between transmitters. At the present time, the petroleum industry has one system installed and working on a 20-kc basis. The equipment is working quite satisfactorily, and additional systems are in the process of being installed on frequencies between existing channels.

Continued expansion of petroleum systems, as well as installations of new systems, can be expected. In order to accommodate these, particularly in the oil-producing areas of the Gulf Coast, both channel-splitting and time-sharing must be utilized. Even so, only the fullest co-operation among users can prevent unbearable degradation of the service.

Radio has truly become one of the indispensable tools in the petroleum industry's never-ceasing struggle to supply the products that make possible our industrial civilization.

◆

No Discussion

Microwave as Applied to Railroad Operation in the Gulf Coast Area

L. R. THOMAS
MEMBER AIEE

THE Santa Fe railroad has installed a microwave system between Galveston and Beaumont, Texas. The installation is being used to supplement existing wire-line communication and to obtain technical information regarding the operation and reliability of a microwave radio-relay system in this area. It was felt at the time the decision was made to install the system that, if microwave radio relay could be made to operate reliably in this area, it could be made to provide reliable service on other parts of the railroad where the terrain and atmospheric conditions would be more favorable to its use.

Fig. 1 shows a map of a part of the Santa Fe's facilities in the Gulf area, indicating that prior to the installation of the microwave system it was necessary for telephone and telegraph circuits to be routed from Galveston to Beaumont, Texas, via Somerville and Silsbee, a distance of 315 miles. From this point they extend on to Longview, Texas, and Oakdale, Louisiana.

The installation of the microwave-relay system on a direct route from Galveston to Beaumont, a distance of approximately 70 miles, makes it possible to pick up the circuits from Somerville east at Beaumont and thus release the physical pairs between Galveston and Somerville for other service. This type of installation also reduces the transmission losses experienced in long physical circuits and provides more satisfactory overall communication from Galveston to the field.

The microwave equipment was supplied by the Philco Company, and consists of their type *CLR-6* terminal and repeaters. There are two terminals and three repeaters in this installation, as shown by Fig. 2, which is an overall diagram of the system showing the antenna heights, distance between repeaters and terminals and the type of antennas used. The equipment operates in the 6,700-megacycle range with the transmitters providing a power output of approximately one watt and receivers adjusted to provide an operating sensitivity of −108 decibels referred to 1 watt.

The multiplexing equipment used with this installation was also supplied by the Philco Company and is pulse-amplitude-modulated equipment capable of providing a total of 24 telephone channels. In this particular installation, equipment to provide only eight telephone channels was used.

Fig. 3 shows the service which is operated over the telephone channels. The first of these is a service channel which is dropped out and inserted at each repeater point to provide communication with field-service personnel. This channel also carries the fault tone signals from the terminal at Beaumont and the three repeaters to the control terminal at Galveston. The second channel is used to pick up the dispatcher's telephone circuit which operates between Somerville, Longview, and Beaumont. The third channel is connected to a party-line telephone to Somerville. The fourth channel is connected to a party-line telephone to Longview. The fifth and sixth channels are used as trunk telephones between Galveston and Beaumont, the seventh is idle, and the eighth carries six channels of frequency-modulated telegraph carrier. This channel is capable of carrying up to 14 channels of telegraph carrier if needed.

The ringing on the trunk circuits is accomplished by means of Western Electric 1000/20-cycle ring oscillators. The

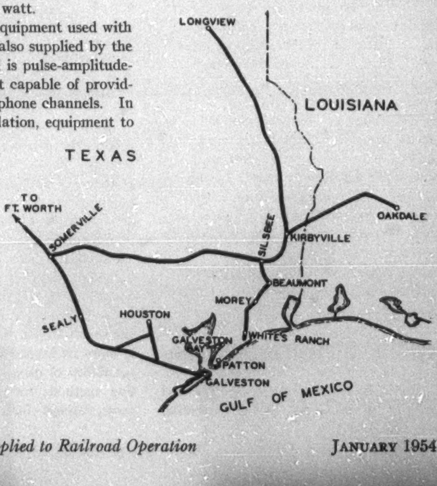


Fig. 1. Map showing a portion of Santa Fe's operation in the Gulf Coast area

---

Paper 53-382, recommended by the AIEE Radio Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Kansas City, Mo., November 2–6, 1953. Manuscript submitted May 14, 1953; made available for printing August 17, 1953.

L. R. Thomas is with The Atchison, Topeka and Santa Fe Railway Company, Chicago, Ill.

The author wishes to acknowledge the assistance of R. B. Grantham and W. H. Veerman, assistant engineers, who performed the field work in connection with the Santa Fe microwave installation.

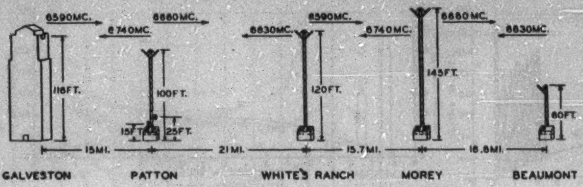

6590MC.
6740MC.
6680MC.
6830MC.
6590MC.
6740MC.
6680MC.
6830MC.
116FT.
100FT.
15FT.
25FT.
120FT.
145FT.
60FT.
15MI.
21MI.
15.7MI.
16.8MI.
GALVESTON
PATTON
WHITE'S RANCH
MOREY
BEAUMONT

Fig. 2. Santa Fe microwave system, Galveston to Beaumont

ringing on the dispatcher's and party-line telephones is $3^1/_2$-cycle selector ringing from Beaumont, controlled from Galveston over telegraph-carrier channels. The ring keys at Galveston operate the send relay on a telegraph-carrier channel, and the receive relay on the telegraph-carrier channel in turn operates the relays of the ring case at Beaumont to provide battery to the physical circuits at that point.

The operators contact the dispatcher by means of his loud speaker in the normal manner. The party-line telephones on the other hand have 20-cycle ringing from the field by means of hand generators. These operate a 20-cycle relay at Beaumont which actuates the send relay of a telegraph carrier channel, and in turn the telegraph-carrier receive relay operates a telephone-type relay at Galveston which applies 20 cycles to the private branch exchange drop at that point. The other telegraph-carrier channels are used either for printer circuit or to pick up Morse wires for termination in the Galveston relay office.

The pulse-amplitude-modulated multiplexing equipment has presented a few problems which can be expected with a new type of equipment, but to date these problems have in most cases been eliminated. However, the signal-to-noise ratio which can be obtained from this type of equipment is not as good as that which can be obtained from land-line carrier circuits. This is to a large extent due to the signal-to-noise ratio of the microwave system itself and is a problem which no doubt will be taken care of in new equipment.

The trouble experienced with the *CLR-6* microwave equipment has reached the stage where it can be said to be of a more or less minor nature. However, it is only fair to point out that the degree of maintenance required to keep a system of this kind in first-class operating condition is rather high. The microwave equipment by its very nature is not a simple device. Therefore, it should be pointed out to those who are contemplating the installation of a microwave system that they should take into consideration the need for more highly trained personnel than is required to maintain a wire-line communication plant. Skilled personnel can be obtained or existing employees trained to satisfactorily maintain the equipment.

When this system was installed it was located in the area where the greatest difficulty could be expected from a standpoint of propagation. Problems were anticipated because of the large bodies of water to be crossed as well as the high moisture content of the air in the Gulf area. Referring to Fig. 4, the path from Beaumont to Morey, a distance of 16.8 miles, consists of relatively dry land with large areas covered with trees and buildings. Very little propagation trouble has been experienced in this area. The exception to this consists of the fades which are encountered when there are rapid changes in temperature or in the moisture content of the air. The fades are normally not of sufficient depth to interrupt service. However, outages do occur at infrequent intervals.

The tower at Beaumont is 60 feet high while the one at Morey is 145 feet. In both cases apparently the trees act as zonal screens which tend to reduce the intensity of fades caused by reflection from the surface of the land between these two points.

Referring to Fig. 5, the section between Morey and White's Ranch, a distance of 15.7 miles, is relatively dry land and there is little if anything in the path except a few trees and ranch buildings. There are also rice farms in this area which are flooded at certain times of the year. The tower at Morey is 145 feet and the tower at White's Ranch is 120 feet. Some rather severe fades have been experienced in this section. The fades are of the type caused by cancellations due to reflection as well as those caused by moisture fronts such as fog banks.

The section between White's Ranch and Patton, Fig. 6, a distance of 21 miles is over rice farms, salt water swamps, and a portion of Galveston Bay. The tower height at White's Ranch is 120 feet and at Patton 100 feet. However, the reflector on top of the tower at Patton is not being used due to severe fading conditions. The parabola employed in this section is located at a point 25 feet above the base of the tower. Comparative tests made with receivers on both the high and the low antenna at Patton indicated such severe fading conditions with the high antennas that satisfactory service could not be obtained. This fading was apparently due to the changing moisture

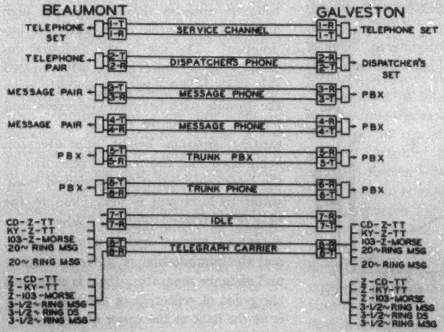
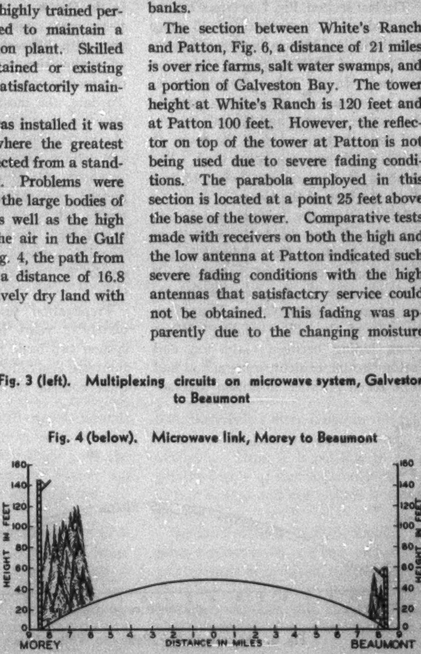

BEAUMONT
GALVESTON
TELEPHONE SET
1-T
1-R
SERVICE CHANNEL
1-R
1-T
TELEPHONE SET
TELEPHONE PAIR
2-T
2-R
DISPATCHERS PHONE
2-R
2-T
DISPATCHER'S SET
MESSAGE PAIR
3-T
3-R
MESSAGE PHONE
3-R
3-T
PBX
MESSAGE PAIR
4-T
4-R
MESSAGE PHONE
4-R
4-T
PBX
PBX
5-T
5-R
TRUNK PBX
5-R
5-T
PBX
PBX
6-T
6-R
TRUNK PHONE
6-R
6-T
PBX
7-T
7-R
IDLE
7-R
7-T
CD-Z-TT
KY-Z-TT
103-Z-MORSE
20~ RING MSG
20~ RING MSG
8-T
8-R
TELEGRAPH CARRIER
8-R
8-T
CD-Z-TT
KY-Z-TT
103-Z-MORSE
20~RING MSG
20~ RING MSG
Z-CD-TT
Z-KY-TT
Z-103-MORSE
3-1/2~RING MSG
3-1/2~RING DS
3-1/2~RING MSG
Z-CD-TT
Z-KY-TT
Z-103-MORSE
3-1/2~RING MSG
3-1/2~RING DS
3-1/2~RING MSG

Fig. 3 (left). Multiplexing circuits on microwave system, Galveston to Beaumont

Fig. 4 (below). Microwave link, Morey to Beaumont



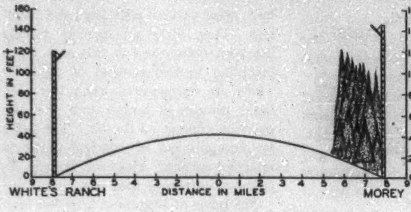


Fig. 5. Microwave link, White's Ranch to Morey

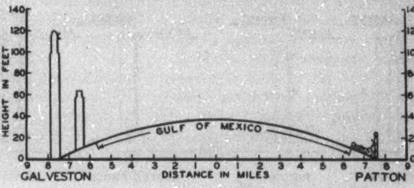


Fig. 7. Microwave link, Galveston to Patton

content of the air over the water and swampy land, as well as some instances of temperature inversion. The characteristics of the fades indicated surface reflection and multipath transmissions resulting in a slight increase in signal, followed by almost a complete cancellation. These fades repeated very rapidly and, in some cases, to the extent that they appeared on the recording voltmeter in a very broad pattern which might be termed "painting." With the low antenna, there was very little fading of this nature. However, there are conditions where the signal gradually decreases to a point of outage, staying out for a period of time dependent upon weather conditions, and then gradually returning to full strength.

The last section, Fig. 7, between Patton and Galveston, a distance of 15 miles over the Gulf of Mexico, has tower heights at Patton of 100 feet and at Galveston 116 feet. In the Patton-Galveston section, the same general condition exists as that between Patton and White's Ranch, and the reflector on top of the tower at Patton is likewise not being used. The parabola is located at a point 15 feet above the base of the tower.

At the time the original installation was made, it was decided to work from parabolas inside of the buildings to passive reflectors on top of towers at all locations, with the exception of Galveston, where the parabola was located on the side of the office building. When severe fading between Galveston and Patton became evident, some action had to be taken in order to obtain satisfactory operation. Several tests were made at Patton with the parabola located at different heights on the tower and recordings were made of the radio-frequency signal strength. Final decision was made to locate the parabola on the tower at a height of 15 feet from the ground, working toward Galveston. This establishes a high-low combination of antennas which changes the point of reflection from the surface of the water approximately midway between the two antennas to a point on the surface of the promontory almost at the base of the Patton tower. This arrangement of antennas practically eliminated the fading problems in this section.

When similar difficulty was experienced in the Patton to White's Ranch section, tests were made over long periods of time with different antenna heights in an effort to achieve the same results as were obtained between Galveston and Patton. To date, the most satisfactory performance has been achieved with the parabola at Patton mounted on the tower at a point approximately 25 feet above the ground. This location gives very satisfactory propagation results during practically all types of weather conditions except fog. During fog conditions, the signal is attenuated to a point where service is disrupted for long periods of time.

Propagation tests covering approximately 2 years indicate that on this system two basic types of fading conditions are generally encountered. One of these consists of a high-maximum low-minimum fade as shown in Fig. 8. This type of fading takes place during periods of temperature inversion or changes in humidity which tend to change the index of refraction, thus changing the path length over which the direct and reflected signals pass. Where the length of the direct signal path in relation to the reflected signal path changes by as much as 1/2 wave length, the receive signals present at the antenna may change from an in-phase to an out-of-phase condition, thereby increasing and decreasing the resultant signal strength. These conditions exist in the Gulf area when winds come down from the North blowing the moisture normally present out to sea, and dropping the temperature sometimes as much as 20 degrees or more in a matter of a few minutes. They also exist during periods where the ambient temperature and the dew-point temperature become the same. Under such conditions condensation takes place and the moisture content of the air increases. This occurs almost every night when there are no storm or fog conditions existing to bring about more violent changes.

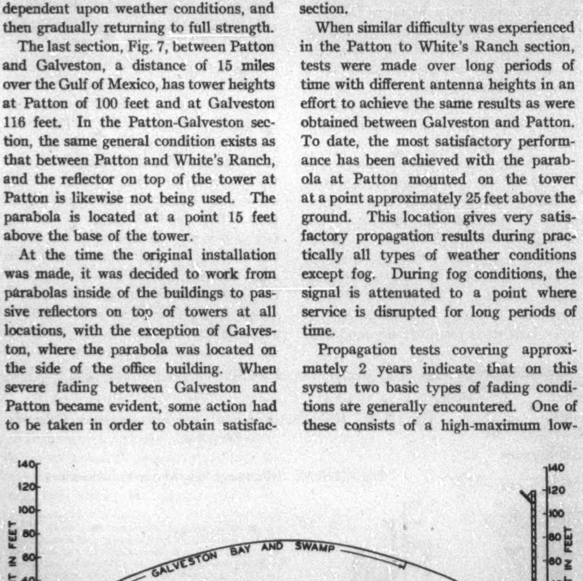


Fig. 6. Microwave link, Patton to White's Ranch

It was noted, with high antennas at both ends of the path, that the high-maximum low-minimum type of fades were more frequent and severe in the Patton to White's Ranch and Galveston to Patton hops than those experienced in other sections. As a result of these tests, it appears that the use of high antennas on paths over water is impractical. The use of the high-low antenna system, however, greatly reduced this type of difficulty, particularly over the 15-mile path between Galveston and Patton. Apparently this is due to the existing antenna height at Galveston being such that the point of reflection of the interfering signal is transferred from water to land and is very close to the base of the tower at Patton. When a high antenna is worked against a low antenna, and the combination of antenna heights and distance separating them is such that the point of reflection of the main beam is at the base of the tower of the low an-

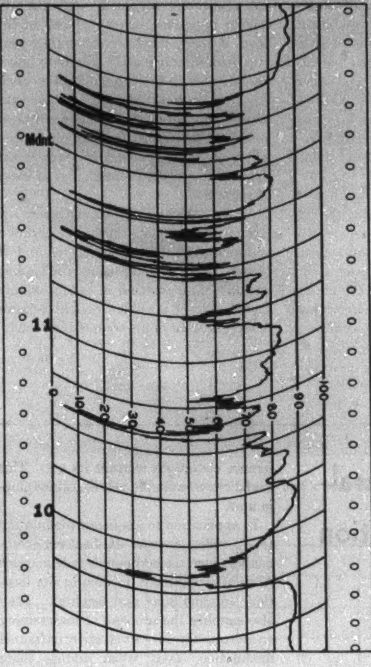


Fig. 8 (left). High-maximum low-minimum type fade

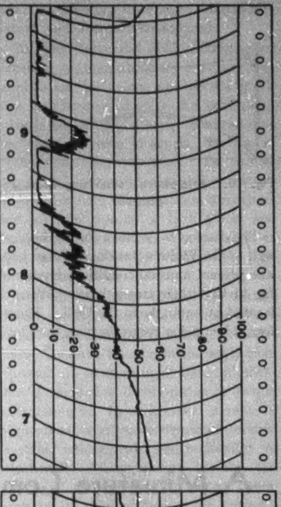


Fig. 9 (right). Fade due to high antenna above fog and low antenna in fog

tenna, the difference in path length is reduced to a value where it is evidently not so important a factor.

In the case of the Patton to White's Ranch hop, Fig. 6, a distance of 21 miles using high-low antennas, the point of reflection is some distance from the base of the tower at Patton, where the low antenna is located. However, due to the prevailing moisture content of the air in this area, the index of refraction is such that the point of reflection is shifted sufficiently to greatly improve the operation, insofar as high-maximum, low-minimum type of fade is concerned.

The second type of basic fade encountered was that brought about by fog. This condition was more prevalent in working a high antenna against a low antenna. During fog conditions, it is not unusual for the signals to be relatively stable and strong between two high antennas, while transmission between a high and low antenna would be attenuated to such an extent as to be unusable. Recent tests conducted revealed a very interesting situation which prevails to some extent in sections where the high-low antenna is used.

Observations were made during the time when fading conditions existed in the early morning hours and the following was determined: When a fog bank starts to move in from the Gulf, whether it is a heavy dense fog or a light ground fog, it is frequently at a low level and, at a later time, drifts upward. When this condition exists a fade occurs, as shown on Fig. 9. This is apparently due to the fact that the low antenna is in the fog layer and the high antenna is above this layer. The transmitted signal from the high antenna to the low antenna hits the surface of the fog layer and apparently a greater portion of the signal is reflected. The amount of reflected signal depends largely upon the density of the fog. This is one of the conditions which gives the greatest trouble between Patton an White's Ranch where the high-low antenna system is used.

During such fog conditions, the Galveston-Patton hop gave a much lesser degree of trouble, but nevertheless a definite reduction in signal level was noticed. Due to the shorter distance between Galveston and Patton and the fact that parabolas are used without reflectors, there is less attenuation in this hop. Therefore, a greater fading margin is available. Thus, it is possible to get a greater degree of signal through the fog in that section and less trouble is experienced.

On the Patton to White's Ranch hop, using a reflector on one end, and with a greater distance involved, the path length provides a lower fading margin. Therefore, on occasions this circuit is completely unusable for several hours at a time.

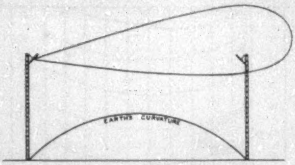


Fig. 10. Microwave transmitting beam adjustment

The Patton to White's Ranch hop, as well as the White's Ranch to Morey hop, also present unusual conditions in case of fog in that their transmission paths are located at an angle to the Gulf. Therefore, when a fog comes in from the Gulf, the front of the fog bank diagonally intersects the transmission path, creating a condition of refraction or reflection that is the same as that produced when one antenna is in the fog belt and the other above. It was noted in both cases that as soon as the fog completely enclosed the antenna on both towers, the signals returned to a relatively high level and stable communication was obtained.

In the same recent tests where it was determined what happened during the fog conditions it was also found that overall stability could be improved if the antennas, Fig. 10, were adjusted so that the half-power point on the bottom of the major lobe of the beam was set on the receiving antenna; that is, only the transmitting antennas were adjusted, and the receiving antennas were so placed in the beam by this adjustment that they were at a point approximately 2 decibels below the midpoint of the beam. Under conditions of normal weather, where the humidity increases slightly between night and day, recordings indicate that there will be a slight increase in level of the signal at night due to the change of index of refraction.

At the present time, arrangements are being made to replace the 4-foot parabolas and reflectors with 6-foot units in this system, thereby increasing the fading margin. This increase in fading margin should make it possible to obtain reliable communication in all hops of the system.

No Discussion

A Miniature Compandor for General Use in Wire and Radio Communication Systems

F. S. BOXALL
NONMEMBER AIEE

R. S. CARUTHERS
MEMBER AIEE

Synopsis: A miniaturized compandor has been developed for widespread application to open-wire, cable, and radio channels. Self-contained plug-in units can be inserted in telephone circuits at all commonly expected transmission levels. Two-wire operation is achieved with a plug-in resistance hybrid. Equipment simplification is achieved through use of only two miniature vacuum tubes. A conventional approach has been followed to obtain a compandor circuit with a two-to-one compression and expansion characteristic. Straightforward empirical methods are used to achieve the desired results. Reduction of noise and crosstalk, commonly stated as 22 decibels (db), actually varies considerably, depending on the type and magnitude of interference.

General

IN RECENT years, a rather remarkable increase has occurred in the use of compandors in the channelizing equipment of speech communication systems. Perhaps even more remarkable is the fact that from an over-all cost standpoint, new carrier installations in which the terminal equipment includes compandors are the least expensive. Compandors have been used for many years to reduce noise and crosstalk on open-wire systems. More recently, in carrier systems for cable and open wire, compandors have been integrated into the design of each channel.[1,2] This has been made possible by the development of a compact, inexpensive compandor which gives quality performance and only slight speech degradation, even when a number of these devices are connected in tandem.

Application

At the Lenkurt Electric Company, objectives in compandor application have included more widespread use on wire and radio systems than previously. Specifically, the aim has been to build a device that can either be used freely or omitted, depending on whether crosstalk or interference conditions warrant its use. This goal dictated a simple, self-contained plug-in unit.

In application to frequency modulation (FM) radio systems at Lenkurt, compandors increase permissible fade margins (thus allowing more freedom in site location, antenna sizes and heights). They also simplify the problem of maintaining an adequately low level of interchannel modulation, even when adding more channels. In another application, a large number of channels is planned for use on a single side-band basis on very high frequency (VHF) or microwave bands. Studies indicate that gains and power levels of the radio frequency amplifiers required are reduced so much by compandor use that a much wider selection of tube types is available.

Probably the most obvious use of a compandor is in the basic design of a carrier system where the use of a compandor in each channel eases all design requirements. The compandor results in more satisfactory modulator levels, lower amplifier gains, and smaller filter discriminations in the terminals. In repeaters and line facilities, at the same

Fig. 1. A shelf holding four compandors is normally mounted in a standard 19-inch telephone type bay

---

Paper 53-384, recommended by the AIEE Wire Communications Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Kansas City, Mo., November 2-6, 1953. Manuscript submitted June 29, 1953; made available for printing August 3, 1953.

F. S. Boxall and R. S. Caruthers are with the Lenkurt Electric Company, San Carlos, Calif.

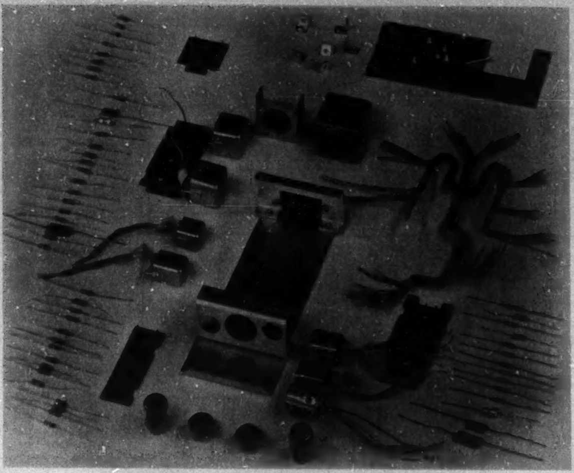


Fig. 2. Compandor frame and components unassembled

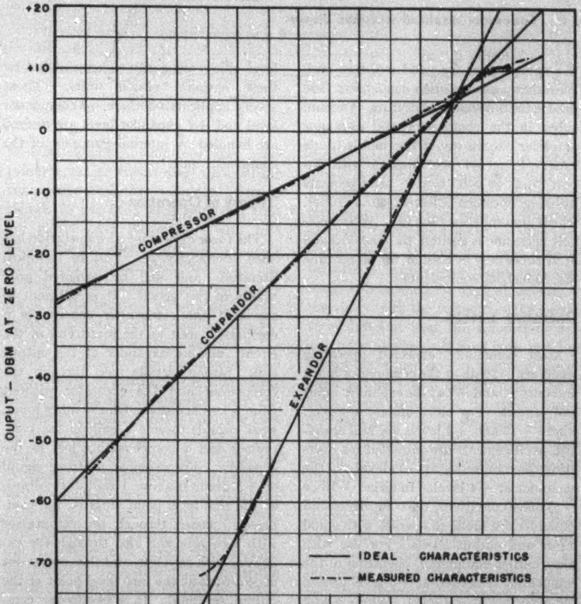


Fig. 3. Compandor load characteristics

time, interference from noise, intermodulation, and crosstalk are reduced.

Another important field of compandor use is in systems where the eventual field application finds some unexpected spurious interference, despite the best planning in the original design. Here, at spot locations in some channels, the use of compandors is warranted. In similar situations, systems designed for a broad geographic field of application sometimes are used in regions of high static and lightning interference. Compandors in such systems are needed only in the high interference areas.

Still another likely field of application is in systems and circuits which may have been installed for many years. Performance may have become increasingly poor through no basic fault of the equipment but through slow deterioration or lack of maintenance. In such cases, it may be difficult even to locate the exact point of system impairment, geographically or electrically. Compandors at the terminals of such circuits will restore operation to its original high-grade performance.

In the railroad and pipe-line field, compandors may be particularly useful in bringing substandard circuits up to high-quality toll performance, or in obtaining satisfactory operation under circumstances which would otherwise provide only intolerably poor circuits. Economies in obtaining circuits of any specified quality may also be effected by the use of compandors.

The compandor described in this paper is not particularly novel from an electrical standpoint, in that it follows closely the design procedures already described in technical literature. However, the actual compandor differs from any previously developed in being smaller, more compact, and in using fewer vacuum tubes. Despite this miniaturization and simplification of circuitry, excellent performance has been retained; and even a skilled observer can find little fault in speech passing through four to eight compandors in tandem.

No attempt has been made to employ other than an orthodox compandor with two-to-one compression and expansion ratios. A conventional syllabic type has been used without exploring the perhaps more intriguing aspects of an instantaneous device. The success achieved, however, indicates that further development might derive even greater noise benefits with higher compression ratios and with further simplification through some compromise between syllabic and instantaneous design. When transistors are

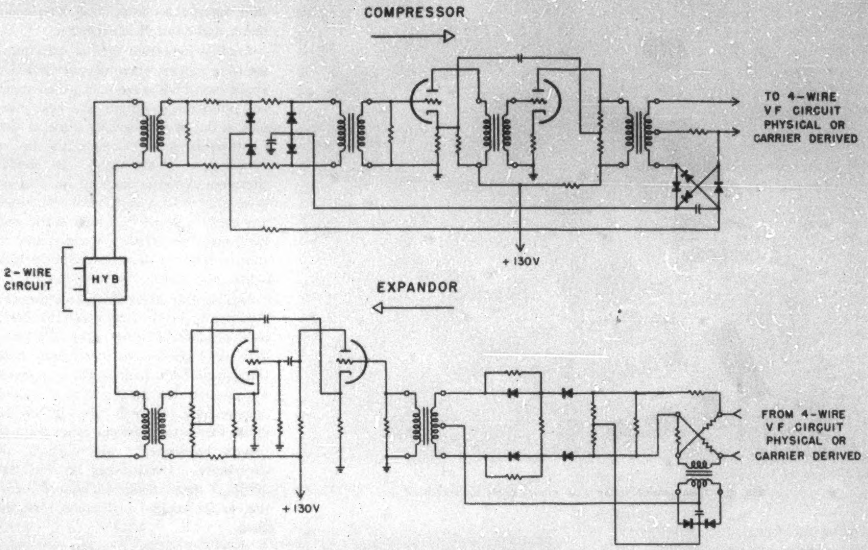


Fig. 4. Compandor simplified schematic diagram

added and cross-control included to permit lower over-all net loss operation, the use of compandors will be certain to expand greatly.

Equipment Arrangements

In Fig. 1, four compandors are shown in a shelf which is normally mounted in a standard 19-inch telephone type bay. One compandor is shown about to be plugged into the framework. Each compandor is approximately 2 inches high, $3^1/_4$ inches wide, and $8^1/_4$ inches deep. External connections are made through a 20-pin multicontact gold-plated plug at the rear of the unit. This plug is inserted in a jack mounted on the shelf. Equipment arrangements are provided for mounting either a single shelf of four units, or twelve units in three decks of four each.

In Fig. 2, an exploded view of one of the units shows the components. In the front are two vacuum tubes, a test socket and a removing handle. To the rear of the unit, an optional plug-in resistance hybrid is shown, permitting 2-wire termination when desired. Within the resistance hybrid is a screwdriver-adjustable potentiometer to set 2-wire net loss. Deposited carbon resistances are used throughout for precision and high stability. Miniature transformers in the amplifier circuits use epoxy resins for good moisture seal. Tantalum electrolytic condensers are used in the time constant circuits. Vacuum tubes in the compressor and expandor amplifier circuits are of the double-triode type. Preferred operation is with 20-volt and 40-volt heater arrangements utilizing Western Electric *407A* tubes, which give superior tube life. Where 6.3-volt operation is desired, the conventional *2C51* tubes may be used or, for better tube life, industrial equivalents.

Operating Levels

Most commonly expected operating levels are −16-db at the compressor input and output, and +7 at the expandor input and output. For older offices, however, where −13 and +4 levels are still standard, arrangements are provided for operating the compressor at −13 level and the expandor at +4 level. In either −16 or −13 office conditions, the compressor and expandor are made to operate with equal input and output levels. For use with some carrier equipment, operation of the compressor is desired at −4 level and of the expandor at +4 level. Where 4-wire operation is desired, the resistance hybrid is removed, and in its place a unit is used with 4-wire connections. Two types are provided to give either zero or 3-db 4-wire loss. Thus, either −16 or −13 level offices are easily accommodated by these optional plug-in units. Those special applications where −4 compressor level and +4 expandor level are desired are handled by internal strapping of the compandor unit.

Theory of Operation

The theory of compandor operation has been thoroughly described in technical literature, and will be reviewed only briefly in this article. It is sufficient to say that the compandor consists of a compressor unit at the input end of the circuit, and an expandor at the output end. Strong signals pass through the compressor and the expandor without gain or loss in either. On the other hand, weak signals receive a gain in the compressor and a compensating loss in the expandor. For example, a strong signal of +5 decibels above 1 milliwatt (dbm), at the zero level point entering the compressor, passes through the compressor with zero gain and also through the expandor with zero loss, and thus comes out at +5 dbm at the zero level point of the output terminal. In a two-to-one compandor, a signal 50 db below a +5 dbm signal receives a 25-db gain in the compressor and a corresponding 25-db loss in the expandor. The compandor noise advantage is derived from the gain given

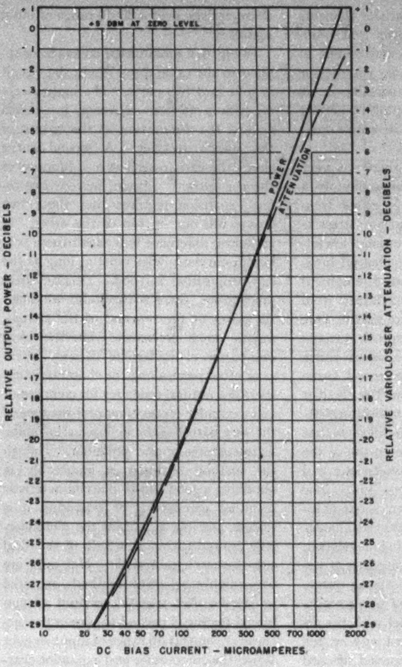


Fig. 5 (left). Compressor d-c operating characteristics

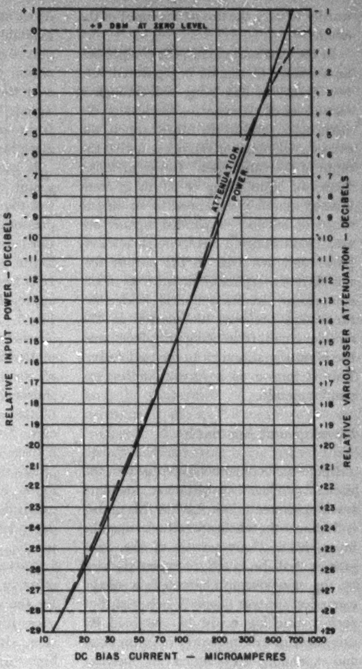


Fig. 6 (right). Expandor d-c operating characteristics

to weak signals so that they may override intervening circuit noise between the compressor and the expandor. Strong signals do not need such gain, and thus are passed over the circuits with zero gain in the compressor and compensating zero loss in the expandor. When the circuit is idle, the expandor retains the loss given to the weakest signal; and a corresponding gain exists in the compressor. This loss during the idle condition is the noise advantage of the compandor, and all interferences existing in the intervening medium between compressor and expandor are subjected to this expandor loss. As speech increases, the compressor gain and expandor loss decrease; while the expandor loss is removed by speech action, thus raising noise, speech masks the increasing noise that results. And thus, even though noise is much increased during the presence of speech, the masking advantage of the speech makes this still a satisfactory operating condition. Time constants of a syllabic length are used in the compressor and expandor circuits. It is important for the time constants to be so arranged that the compressor gain and expandor loss is reduced quickly upon a speech burst. At the end of a speech burst, it is likewise important that the compressor gain and expandor loss slowly recover to their idle values. No great precision is required in these time constant values. In fact, it is even difficult to state definitely what these time constant values are in a typical circuit, as they vary with the strength of the signal. It is sufficient that the compressor and expandor attain their final gain and loss values upon insertion of a strong speech burst within 1 to 5 milliseconds. Similarly, upon removal of the speech burst, restoration of the compressor and expandor gains to their original values within 20 to 100 milliseconds is probably adequate.

In Fig. 3, input-output curves are shown for compressor and expandor. Both ideal and measured curves are shown. Compressor and expandor curves are plotted as though each were operated at zero level at its input and output branches instead of the actual −16 and +7 levels respectively. An input of +5 dbm at zero level receives zero gain in both compressor and expandor. A 60-db weaker input to the compressor receives 30-db gain in the compressor and 30-db loss in the expandor. Failure of the compressor and expandor to introduce equal gain and loss introduces an over-all circuit error. With the compandor shown, this error is less than 1 db over a range of input signals of nearly 60 db.

Circuit Arrangement

In Fig. 4, the compressor and expandor circuit arrangements are shown. In the compressor circuit, a range of signals of approximately 60 db from the resistance hybrid or 4-wire input is compressed to half this amount by the variolosser circuit, and then fed through a 2-stage feedback amplifier to the output. At the output this compressed range of signals is fed through a rectifier circuit to control the loss of the variolosser at the input end of the circuit. At the output terminals, a coil hybrid arrangement isolates the output signal branch from the rectifier branch to prevent harmonic distortion of the rectifier from being fed into the output signal branch. The arrangement shown is known as a backward-acting device. The compressed output band acts backwardly to vary the loss at the input of the circuit.

In the expandor circuit, the compressed input range of signals is separated into two branches, one being fed through a rectifier circuit to control the loss of the variolosser; and in the other branch, the compressed range of signals is fed to the input of the variolosser. In this arrangement, an input range of 30 db of compressed signals is expanded to an output range of 60 db. Expanded signals are amplified through a 2-stage feedback amplifier and delivered to the output terminals of the expandor circuit. As in the compressor circuit, a hybrid bridge arrangement prevents modulation in the rectifier circuit from being introduced into the input signal branch to the variolosser. This is known as a forward-acting expandor circuit.

Development Approach

In the development of a compandor, the greatest problem is to obtain an adequate tracking characteristic. What this means is that over the whole operating range, the input-output characteristic of the compressor shall be exactly compensated for by the input-output characteristic of the expandor. It was formerly believed that perfect tracking could be obtained by using true inverse networks in compressor and expandor. In practice this does not work out, and the approach must be much more empirical. This is because of limitations imposed by circuit operating levels and limitations of varistor elements available. Experience shows that probably the best approach is to attempt to obtain a perfect two-to-one compressor over the entire range of operating input signals and, in turn, to obtain a perfect two-to-one expandor over its entire operating range.

It has been found that there is a very useful approach in obtaining directly a two-to-one characteristic in either compressor or expandor without any stress as to whether varistor characteristics are following square laws or exponential laws. It is merely a straightforward, empirical approach. This can be explained from Fig. 5. Here two curves are shown. One shows the rectified output speech currents plotted as a function of the output range of compressor signals. A second curve shows the d-c current from the rectifier plotted versus loss in the variolosser circuit. A simple examination will show that if these two curves fall one on top of the other, the compressor characteristic will follow exactly a two-to-one law. The experimental approach, then, is one of independently measuring these two characteristics in the circuit being developed, and attempting to move one curve or the other to fall on top of the second curve. By the simple expedient of using resistances in series or in parallel with the diodes, either the low-current end or the high-current end may be moved upward or downward, as desired. The curves in Fig. 5 show how closely this empirical approach was able to obtain a two-to-one characteristic over a 60-db range of input signals.

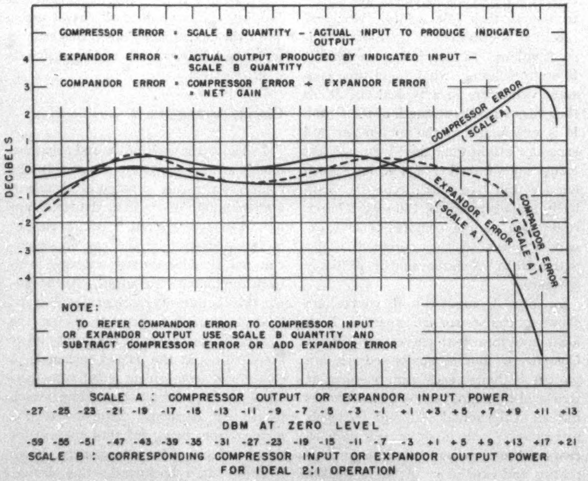


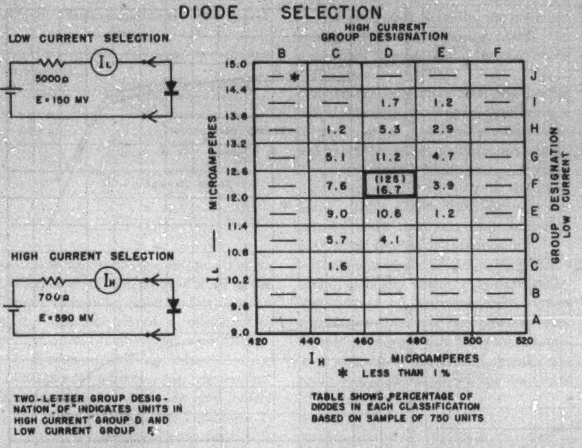
Fig. 7. Compandor error characteristics

In Fig. 6, a similar method has been applied to the expandor. Again the circuit is split into two parts. The input range of signals to the expandor is plotted against the d-c current at the output of the speech rectifier. A second curve shows the variolosser loss as a function of the d-c current. Again, perfect two-to-one expansion requires that these two curves fall one on top of the other. It makes no difference how much these particular curves bow or twist as long as the superimposition is good. Despite these precautions, there is inevitably at either the lower or the top end of the characteristic some departure from the ideal two-to-one characteristic desired. Here as at other parts along the characteristics, there is a natural tendency for compressor and expandor errors to cancel because of the very nature of the inverse action going on in compressor and expandor. Though the method outlined is invaluable in obtaining a good two-to-one characteristic in either compressor or expandor, it is seldom possible to obtain the characteristic desired within 1 or 2 db of the ideal two-to-one characteristic. Not only are very highly selective methods required for diodes, but a great amount of empirical testing is needed to obtain this accuracy over a 60-db range of input signals. When the compressor and expandor characteristics are added, these deviations can be reduced to the order of magnitude of about ±0.5 db with good design.

In Fig. 7, deviation errors are plotted for compressor, expandor, and over-all circuit. The tendency of compressor and expandor errors to compensate is seen.

Selection of Diodes

Obtaining a desired compressor or expandor characteristic is, of course, fully dependent on obtaining germanium diodes with the proper characteristics. In using the procedure just described to obtain good tracking, it was found that many sources of supply of suitable diodes were available. The procedure was to classify a large number of diodes from a particular manufacturer with generally suitable characteristics, find the highest concentration of diode characteristics in this lot, and then make the basic compandor circuit design suitable for the high concentration area of diodes. In exploring and classifying diodes, 2-point current measurements were used. Diodes were first grouped by resistance at about 500 microamperes (μa), and also at a current in the neighborhood of 12 μa, using a cir-



	B	C	D	E	F
J	— *	—	—	—	—
I	—	—	1.7	1.2	—
H	—	1.2	5.3	2.9	—
G	—	5.1	11.2	4.7	—
F	—	7.6	(125) 16.7	3.9	—
E	—	9.0	10.6	1.2	—
D	—	5.7	4.1	—	—
C	—	1.6	—	—	—
B	—	—	—	—	—
A	—	—	—	—	—

TWO-LETTER GROUP DESIGNATION "DF" INDICATES UNITS IN HIGH CURRENT GROUP D AND LOW CURRENT GROUP F.

TABLE SHOWS PERCENTAGE OF DIODES IN EACH CLASSIFICATION BASED ON SAMPLE OF 750 UNITS

Fig. 8. Diode selection

cuit shown in Fig. 8. Diodes were grouped at high current into five classes: *B*, *C*, *D*, *E* and *F*, using current steps of 20 μa from 420 to 520 μa. At low currents, diodes were grouped into 10 classes, ranging from 9 to 15 μa in steps of 0.6 μa. Particular diodes were given a 2-letter designation, thus, a *BA* diode had a current where *B* designated the high-current value and *A* designated the low-current value. A *BA* diode had a high-current value of between 420 and 440 μa, and a low-current value between 9 and 9.6 μa. A distribution of 750 diodes classified by this method is shown in Figure 8. The largest concentration occurred in the *DF* category, which included about 1/6 of the diodes tested. Consequently, the design of the compressor and expandor uses diodes only in this category. The 750 diodes studied were received from the manufacturer in three separate lots. In each instance, the *DF* category showed largest concentration, giving some indication that successive samples would not show wide departure from this type of distribution.

In Fig. 9, current-resistance plots are shown for the diodes just described. The ranges of resistance at high-current value are shown as well as those for low-current values. With such tight control limits specified for the *DF* units, the log-log current-resistance plot is practically certo be a straight line.

Frequency Characteristic

The frequency characteristics of the compandor are shown in Fig. 10. The over-all characteristic was made by the normal procedure of varying input frequencies to the compressor and measuring the levels of frequencies at the expandor output. In actual use, the frequency characteristic for speech is not determined quite in this manner. What happens is that the syllabic energy with its heavy concentration near the 1,000-cycle point sets the loss of the expandor and the compressor for the syllabic interval. During this period, nearly constant current flows in the compressor and expandor speech rectifier control circuit. The variolosser networks then act as fixed resistance pads, and there is no tendency for compressor or expandor variolosser circuits to double or halve errors that might ensue in the measurement with single frequency. Actually, in practice the frequency characteristics of control circuits have been carefully designed so that there is little difference between an actual over-all single-frequency measurement and the action with speech. The frequency characteristics shown were measured both with fixed bias simulating speech action, and with bias resulting from the input frequencies. There is little difference indicated in the test results.

In Fig. 11, harmonic distortion is plotted for the over-all compandor. In both the compressor and expandor, it was necessary to use special means to prevent harmonic distortion produced in the speech rectifiers from being reflected into the output signal branch. This separation was obtained by use of a hybrid bridge at the output of the compressor circuit. In the expandor, a similar hybrid used at the input where the speech was separated off for the power rectifier circuit and for the branch feeding to the variolosser, prevented reflection of harmonic distortion from the speech rectifier into the variolosser. It was assumed that in some cases there would be practically 100 per cent reflection from the input impedance feeding the expandor input

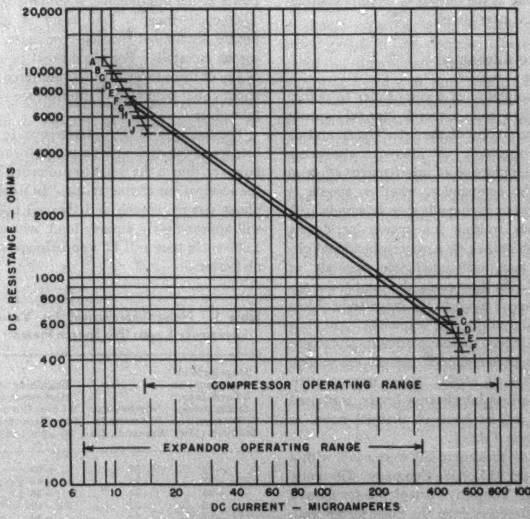
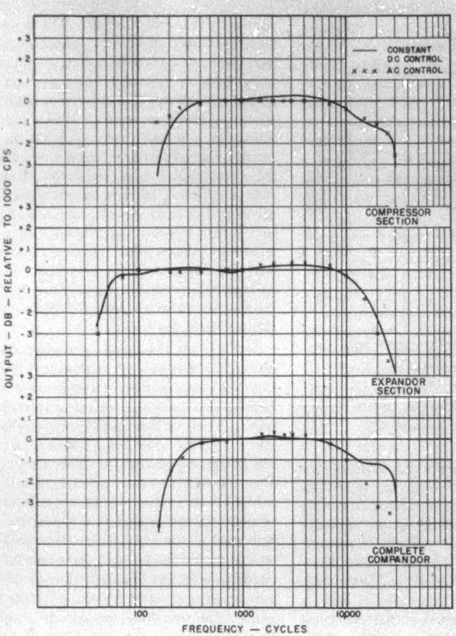


Fig. 9. Diode operating characteristics



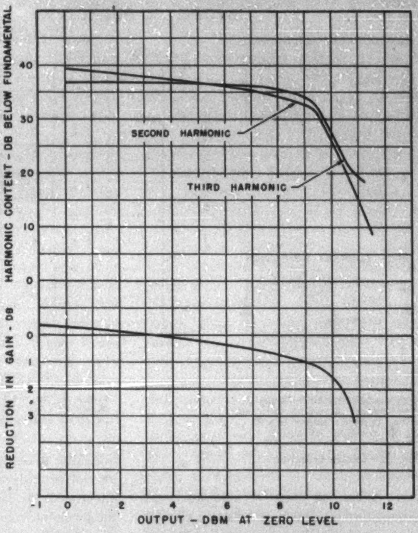


Fig. 10 (left). Compandor frequency response characteristics

Fig. 11 (above). Compandor overload and distortion

terminal. This is most likely to occur above the passed band when expandors are put at the output of circuits containing low-pass filters.

Noise Advantage

The ability of a compandor to reduce interference in a speech channel depends on the nature of the interference and on its magnitude. A difference also exists in the reduction of interferences afforded by the compandor, whether speech is present or during idle speech periods. A listener waiting for answer or during speech pauses on a noncompandored circuit may hear intelligible crosstalk or tones. In these cases, the compandor advantage is equal to the magnitude of loss introduced by the expandor. At low noise levels, the noise advantage exceeds 30 db, becoming progressively less at high interference levels. Noise improvement for various interference levels, measured with a 3-kc band of thermal noise, is shown in Table I.

These measurements were made with a Western Electric Company *2B* noise meter. Subjective listening tests showed no differences from the noise meter readings.

The interference reduction in the expandor will depend on the magnitude of power in the interference, whether it be a 1,000-cycle tone, a 3-kc band of uniformly distributed thermal-type noise, or speech crosstalk. For equal input power of any of these three types of interference, about equal reduction of interference will take place.

The loudness of the interference to a listener at the expandor output will vary, however, due to the weighting afforded by the receiver ear characteristic. In a telephone circuit, speech and thermal noise will appear nearly equally loud, whereas 1,000-cycle tone will be approximately 8 db louder.

Table I. Noise Improvement for Various Interference Levels (No Speech Present)

Magnitude of Interference at 0 Level Input to Expandor, Db from 1 Milliwatt (Mw)	Interference Reduction in Expandor, Db	Magnitude of Interference at 0 Level Output from Expandor, Db from 1 Mw
−32	37	−69
−27	32	−59
−25	31	−56
−22	27	−49
−17	22	−39
−12	17.5	−29.5

In the case of intelligible crosstalk, with crosstalk intelligible only during absence of speech on the listener's channel, the expandor advantage appears in Table I. However, because of the compressor action in raising signal level, a crosstalk disadvantage will occur in the disturbing talker's compressor circuit. The magnitude of this disadvantage as a function of the talker volume is shown in Table II.

The talker volumes shown cover practically the entire range expected at the input to a toll telephone circuit. Since practical systems must be designed to tolerate crosstalk from the louder talkers, the compressor disadvantage is slight. A statistical study will show that a compressor disadvantage of about 5 db, corresponding closely to a −5-volume-unit (vu) volume talker, may be used for computing intelligible crosstalk. As an example, in a line with 25-db crosstalk coupling, a −5-vu talker would be raised

Table II. Crosstalk Disadvantage Resulting from Compressor Action

Talker Volume at 0 Level, Vu	Compressor Disadvantage, Db
−30	17.5
−16	10.5
− 5	5.0
0	2.5

Table III. Noise Improvement with Speech Present

Magnitude of 1,000-Cycle Interference Dbm at 0 Level	Compandor Noise Advantage Speech Volume at 0 Level −26 Vu	−16 Vu	0 Vu
−30	27	25	24
−25	24	22	20
−20	22	19	18
−15	18	16.5	15
−10		14	11

to 0 vu at the compressor output. The interference would appear at −25 vu at the listener's expandor input and at −56 vu at the expandor output, a crosstalk coupling of 51 db (31-db expandor advantage; −5-db compressor disadvantage).

Compandor advantage in noise reduction during speech has been very loosely quoted as 22 db. Actually, the noise advantage varies considerably with the strength of the speech and the magnitude of the noise.

Extensive tests have been made using both 1,000-cycle tone and a 3-kc band of thermal noise as interference. Three typical expected talker volumes were used: 0 vu representing maximum, −16 vu as an average, and −26 vu near the minimum. Table III shows compandor advantage under these test conditions using 1,000-cycle tone as disturbance.

These test results are averages made by careful listeners whose judgment generally differed by less than 1 db. Most spread was observed at greatest compandor advantage and least interference level where tone from the compandor was near threshold of audibility for some listeners.

Observers had much more difficulty in evaluating compandor advantage when a 3-kc band of thermal noise was used in tests. However, by repeated tests it was evident that individual observers straddled the mean values obtained in 1,000-cycle measurements. It is probably a safe conclusion that no difference actually exists as far as the ear is concerned.

References

1. THE N-1 TELEPHONE CARRIER SYSTEM: OBJECTIVES AND TRANSMISSION CHARACTERISTICS, R. S. Caruthers. *Bell System Technical Journal*, vol. 29, no. 1, New York, N. Y., Jan. 1951, pp. 1–32.

2. TYPE-O CARRIER TELEPHONE, J. A. Coy, E. K. Van Tassel. *AIEE Transactions*, vol. 72, pt. I, 1952 (Jan. 1953 section), pp. 428–37.

◆

No Discussion

The Solution of Sturm-Liouville Problems by D-C Network Analyzer

GEORGE W. SWENSON, JR.
ASSOCIATE MEMBER AIEE

A GENERAL second-order ordinary differential equation occurring frequently in engineering analysis is

$$\frac{d}{dx}\left[f(x)\frac{d\phi}{dx}\right]+\lambda g(x)\phi=0 \qquad (1)$$

in which λ is a constant and f and g are arbitrary functions such that a finite solution exists in the interval of interest. It is the purpose of this paper to show how solutions satisfying this equation and various boundary conditions can be obtained by means of a simple d-c network analyzer. For such an equation, there exist only certain values of the parameter λ, the eigenvalues, which permit the solution to satisfy the boundary conditions. The first problem, then, is to determine one or more of these eigenvalues, whereupon the corresponding solutions can be obtained. This procedure presents considerable difficulty, and analytical solutions are available only for a few special cases.

Numerical solutions can be obtained, at least in principle, with the aid of some results from the calculus of finite differences.[1] Consider Fig. 1, which represents an interval along the x-axis, divided into a number of equal segments of length $h/2$. If h is very small, the following expressions are approximately true:

$$\left(\frac{d\phi}{dx}\right)_0=\frac{\phi_b-\phi_a}{h}$$

$$\left(\frac{d^2\phi}{dx^2}\right)_0=\frac{\phi_1+\phi_2-2\phi_0}{h^2}$$

$$\left(\frac{df}{dx}\right)_0=\frac{f_b-f_a}{h}$$

$$f_a=\frac{f_0+f_1}{2} \qquad (2)$$

$$f_b=\frac{f_0+f_2}{2}$$

$$\left(\frac{df}{dx}\right)_0=\frac{f_2-f_1}{2h}$$

$$\left(\frac{d\phi}{dx}\right)_0=\frac{\phi_2-\phi_1}{2h}$$

Equation 1 can be written

$$f\frac{d^2\phi}{dx^2}+\frac{df}{dx}\frac{d\phi}{dx}+\lambda g\phi=0 \qquad (3)$$

Substitution into equation 3 from 2 yields a finite-difference expression, which holds at point 0

$$f_0\frac{\phi_1+\phi_2-2\phi_0}{h^2}+\frac{f_2-f_1}{2h}\,\frac{\phi_2-\phi_1}{2h}+\lambda g_0\phi_0=0$$

or

$$(4f_0-f_2+f_1)\phi_1+(4f_0+f_2-f_1)\phi_2-4(2f_0-\lambda g_0h^2)\phi_0=0 \qquad (4)$$

A similar equation holds at each numbered point in the domain; therefore, given an appropriate value of λ, and values of f and g at each point, an approximate numerical solution of the problem can be obtained by solving the several difference equations simultaneously. As the set of equations is homogeneous, it is necessary that their determinant be zero in order that a nontrivial solution exist. By extracting the zeros of the determinant, therefore, eigenvalues of λ in number equal to the order of the determinant can be determined. However, if a large number of eigenvalues is needed, or if considerable accuracy is needed in the lowest few, it is necessary to use a large number of lattice-points and, thus, a determinant of high order. Because of the great labor involved in extracting the zeros of such a determinant, and in solving the set of homogeneous equations, it

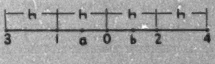


Fig. 1. Finite-difference lattice

Paper 53-412, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Kansas City, Mo., November 2–6, 1953. Manuscript submitted January 28, 1953; made available for printing August 21, 1953.

GEORGE W. SWENSON, JR., is with Washington University, St. Louis, Mo.

The author wishes to express his appreciation to Professors H. A. Peterson, T. J. Higgins, and K. F. Wendt, of the University of Wisconsin, for the loan of equipment belonging to that institution.

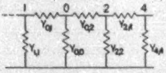


Fig. 2. The electric network

is usually necessary to use some numerical or experimental method to effect the solution.

The D-C Network Analyzer

A set of linear algebraic equations of the form of equation 4 can be used to describe the junction-point currents of a network of linear electric-circuit elements. The network takes the form of Fig. 2, in which the node-to-ground voltages correspond to the ϕ's of equation 4. The admittances $Y_{0,2}$ and $Y_{0,1}$ are equal to the coefficients of ϕ_2 and ϕ_1, respectively, and the admittance $Y_{0,0}$ is such that the sum of the admittances connected to point 0 equals the coefficient of ϕ_0.

It may occur that both positive and negative admittances are required. In this event, the use of inductors and capacitors and alternating currents suggests itself. However, substantial gains in economy and accuracy can be accomplished by the use of resistors and direct currents only, the required "negative resistors" being simulated by active networks containing resistors and voltage sources. The negative-resistor circuit is illustrated in Fig. 3, in which, when the voltmeter is made to read zero by manipulation of R_s, a current of magnitude V_n/R_n exists in the external circuit, of such polarity that the input resistance is negative and equal in magnitude to R_n.

If equation 1 is to be solved in its most general form, it is necessary so to construct the analyzer that both node-to node and node-to-ground admittance be adjustable through wide ranges of positive and negative values. This, in turn, requires that a separate battery be provided for each negative resistor. A considerable simplification is achieved if $f(x)$ in equation 1 is a constant. In this case, equation 1 becomes

$$\frac{d^2\phi}{dx^2}+\lambda p(x)\phi=0 \tag{5}$$

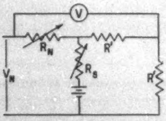


Fig. 3. The negative-resistor circuit

where $p=g/f$. The finite-difference equivalent applying at point 0 becomes, after multiplying by -1

$$-\phi_1-\phi_2+(2-\lambda p_0h^2)\phi_0=0 \tag{6}$$

Similar equations can be written for the other points in the interval

Point 1: $-\phi_3-\phi_0+(2-\lambda p_1h^2)\phi_1=0$

Point 2: $-\phi_4-\phi_0+(2-\lambda p_2h^2)\phi_2=0$, etc.

The network for this simpler case is the same as that shown in Fig. 2, except that here the node-to-node resistances are all of the same positive value. It is clear, however, that the node-to-ground resistances may still be negative; in fact, if p is positive, any point n, whose neighboring ϕ's are nonzero, must have a node-to-ground resistance of $-R/(\lambda p_nh^2)$ ohms, where R is the common node-to-node resistance. All of these negative resistors have a common ground terminal; therefore, only one voltage source need be used for all of them. Such an analyzer[2] can be constructed with the node-to-node resistors, and the resistors R' of the negative-resistance circuits permanently wired with jacks provided for plugging in resistors R_s and R_n of the negative resistance circuits, or the positive node-to-ground resistors, whichever are needed. Jacks are also provided so that one voltmeter can be used in the adjustment of all circuits.

The setup and operation of such an analyzer, as well as methods of accommodating various boundary conditions, will be illustrated in connection with a torsional vibration problem. A second example will illustrate a more general problem.

Example 1. Torsional Vibration of a Uniform Shaft with a Flywheel

Consider a uniform shaft of length L, torsional rigidity C, and polar moment of inertia I per unit length, fixed at one end and free at the other, with a flywheel of moment of inertia $7IL/10$ fixed to the free end. Let it be desired to determine the natural frequency and eigenfunction corresponding to the lowest mode of torsional oscillation. The equation of motion, for sinusoidal oscillations, is

$$\frac{d^2\theta}{dx^2}+\frac{\omega^2h^2I}{C}\theta=0 \tag{7}$$

where θ is the angle of twist of the shaft, and the boundary conditions are

$$\theta=0 \text{ at the fixed end} \tag{8}$$

$$C\frac{d\theta}{dx}=7\omega^2IL/10 \text{ at the flywheel end} \tag{9}$$

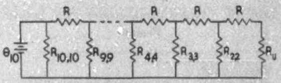


Fig. 4. Analyzer setup for a torsional vibration problem

If the interval L is divided into ten equal segments h, equation 7 can be represented approximately by nine difference equations, one for each interior lattice-point, numbered consecutively from the fixed end to the flywheel end. The boundary condition at the flywheel adds a tenth difference equation to the set, which is then

$$\theta_1(2-\omega^2h^2I/C)-\theta_2=0 \tag{10.1}$$

$$-\theta_1+\theta_2(2-\omega^2h^2I/C)-\theta_3=0 \tag{10.2}$$

$$\vdots$$

$$-\theta_8+\theta_9(2-\omega^2h^2I/C)-\theta_{10}=0 \tag{10.3}$$

$$-\theta_9+\theta_{10}(1-7\omega^2h^2I/C)=0 \tag{10.4}$$

Note that equation 10.1 contains only two terms, one of the neighboring points to point 1 having zero displacement. This is one boundary condition, the other being represented by equation 10.10. The analyzer setup is shown schematically in Fig. 4, in which $R_{1,1}=R/(1-\omega^2h^2I/C)$, $R_{2,2}=R_{3,3}=. \quad . \quad .=R_{0,9}=-R/(\omega^2h^2I/C)$, and $R_{10,10}=-R/(7\omega^2h^2I/C)$. The battery θ_{10} establishes the amplitude of vibration of the flywheel, and is necessary in order to obtain a nontrivial solution.

After the appropriate resistors have been plugged into the analyzer, the resistors R_s are adjusted in sequence to make all voltages V equal zero. This is an iterative adjustment, as each change affects all the other negative resistors; however, convergence is reasonably rapid in most cases. So far, nothing has been said about the determination of the natural frequencies. It is necessary to choose an arbitrary value of ω for the initial setting up of the analyzer, but the addition of the battery $\theta_{1,0}$ makes the system of nodal equations inhomogeneous, so that a solution is obtained for any ω. The battery current corresponds to a torque applied to the flywheel, and the lowest natural frequency can be found by trying different values of ω until the battery current is zero. This indicates that the system is oscillating without the application of any external torque; that is, oscillating in a normal mode. In this manner the lowest frequency of the system under discussion was determined, along with its eigenfunction, in about one hour's time, not including the time required for preliminary computations

The value obtained was $\omega = 0.976CI/L$, which is within 0.3 per cent of the analytically computed value. The analyzer components were calibrated on a Wheatstone bridge to within 1 per cent of their desired values, the forcing battery θ_{10} was $1^1/_2$ volts, and all voltages V were adjusted to less than 1 millivolt.

The iterative-adjustment process does not converge for higher modes; in fact, it does not converge for frequencies higher than the lowest at which the mechanical driving-point impedance becomes infinite. Another scheme must be used, therefore, for determining higher natural frequencies and eigenfunctions. For the second natural frequency, an estimate is made as to the position of the nodal point on the shaft. Then the lowest natural frequencies of the two portions of the shaft, as separated by the nodal point, are found and compared. If they are not equal, the nodal point is moved and the process repeated. Thus, by trial and error, the second frequency can be found. A similar, but considerably more tedious, process permits the determination of the third and higher modes.

Example 2. Torsional Vibration of a Nonuniform Shaft

Suppose that the shaft of example 1 were of two different diameters, the discontinuity occurring between points 4 and 5. Let the torsional rigidity be $2C$ between the fixed end and 4, and C for the rest of the shaft. Similarly, let the two polar moments be I_a and I_b, respectively. The nine difference equations corresponding to equation 4 can now be written, as well as a tenth equation expressing the boundary condition at the flywheel. The coefficient matrix is as follows

$$\begin{bmatrix} B_1 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -2 & B_2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & B_3 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -9 & B_4 & -7 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -5 & B_5 & -3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & B_6 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & B_7 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & B_8 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & B_9 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & B_{10} \end{bmatrix}$$

where

$$B_1 = B_2 = B_3 = (4C - \omega^2 h^2 I_a)/C$$
$$B_4 = 4(4C - \omega^2 h^2 I_a)/C$$
$$B_5 = 4(2C - \omega^2 h^2 I_b)/C$$
$$B_6 = B_7 = B_8 = B_9 = (2C - \omega^2 h^2 I_b)/C$$
$$B_{10} = (1 - 7\omega^2 h^2 I_b)/C$$

Note that the matrix is assymmetrical, and that the equations thus have no analogous network of bilateral resistors. This situation can be remedied by multiplying the fourth equation by 2/9, the fifth by 14/45, and each succeeding one by 14/15. The resulting system can be set up on an analyzer similar to the one in the preceding example, except that all node-to-node resistors must now be capable of adjustment. Negative resistors are needed only in node-to-ground positions, as before.

Forced Oscillations

If a function of x is added to the right-hand side of equation 1 the equation becomes inhomogeneous, and a solution is possible for any value of λ. This corresponds to forced simple-harmonic vibration of a shaft, for example, and can be simulated by inserting appropriate currents into the various nodes of the network analyzer. Thus, problems in forced vibrations of distributed-parameter systems can be solved quite conveniently on the analyzer.

References

1. Numerical Methods in Engineering (book), M. G. Salvadori, M. L. Baron. Prentice-Hall, Inc., New York, N. Y., 1952.

2. A Direct-Current Network Analyzer for Solving Wave-Equation Boundary-Value Problems, G. W. Swenson, Jr., T. J. Higgins. *Journal of Applied Physics*, New York, N. Y., vol. 23, Jan. 1952, pp. 126–30.

3. The Solution of Waveguide and Cavity Resonator Problems With the Resistance-Network Analog, G. Liebmann. *Monograph No. 38*, The Institution of Electrical Engineers, London, England, May, 1952.

4. Analysis of Nonuniform Columns and Beams by a Simple D. C. Network Analyzer, G. W. Swenson, Jr. *Journal of Aeronautical Sciences*, New York, N. Y., vol. 19, 1952, pp. 273–76.

◆

No Discussion

Solutions of Boundary-Value Problems on Automatic Computing Equipment

FRANK M. VERZUH
ASSOCIATE MEMBER AIEE

THIS investigation is primarily concerned with the solution of boundary-value problems by means of automatic computing equipment. Specifically, it is concerned with the solution of ordinary differential equations in which certain conditions (boundary values) are specified at two or more values of the independent variable. Such problems are generally called boundary-value problems; however, it must be emphasized that this report is restricted to the solution of ordinary differential equations, and not to the solution of partial differential equations.

It is clearly evident that the solution of boundary-value problems is inherently more difficult than the solution of initial-value problems, in which all conditions are specified at a single value of the independent variable. The particular problems studied are further complicated by the fact that their solutions are required over an infinite range of the independent variable, a range which contains regions of nonanalyticity. The presence of such singularities materially complicates the solution of the problem.

The following digital computers were used in this study:

1. Conventional 10-digit desk calculator;
2. Type-*602A* Electromechanical Calculator;
3. Type-*604* Electronic Calculator;
4. Card-Programmed Calculator;
5. Massachusetts Institute of Technology (MIT) Whirlwind Computer.

In addition analogue studies were obtained by means of the MIT Rockefeller Differential Analyzer.

There are at least four methods for solving the boundary-value problem:

1. Direct algebraic method—solution of simultaneous algebraic equations;

Paper 53-395, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Kansas City, Mo., November 1–4, 1953. Manuscript submitted March 18, 1953; made available for printing August 13, 1953.

Frank M. Verzuh is with the Massachusetts Institute of Technology, Cambridge, Mass.

This paper is based on some of the research investigations which were performed as part of a Doctorate Thesis at the Massachusetts Institute of Technology.

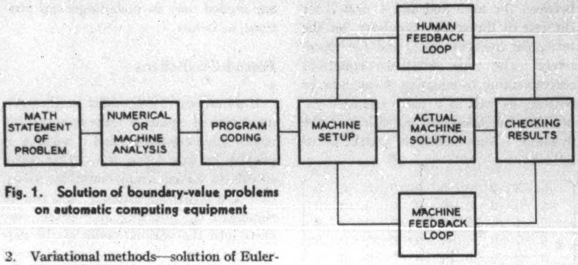


Fig. 1. Solution of boundary-value problems on automatic computing equipment

2. Variational methods—solution of Euler-Lagrange equations;

3. Integral-equation method—solution of the integral equation by integration;

4. Trial-and-error method—successive solution of the initial-value problems.

In the course of this research all of these methods were employed with a varying degree of success. The trial-and-error method of solution was found to be most amenable to solution by means of automatic computing equipment, and all succeeding comments will be concerned with this method of solution of the boundary-value problem.

The utilization of digital computing equipment for the solution of the boundary-value problem naturally raises the following questions:

1. How many decimal digits D must be carried during the computation to insure an accuracy of S significant digits in the final results?

2. Since the inherent use of an approximation formula introduces a truncation error T in the computation, what type of approximation formula is most suitable for the computation?

3. The choice of the interval of integration h strongly influences the round-off error incurred in the computation as well as the truncation error. What is the optimum interval of integration to be used to insure a maximum accuracy of computation?

4. How is the accuracy of computation S influenced by the presence of singularities in the problem?

5. Is the proposed computational procedure amenable to solution by the available automatic computing equipment?

6. Is the automatic machine solution justified from the standpoint of sound economic and engineering considerations?

This investigation provided definite answers to these questions. Obviously, these answers are influenced by the type of computer used in the computations as well as by the method of solution. The next section contains a computational philosophy which describes the basic elements involved in the computational procedure.

Comments on Computational Philosophy

The effective mechanization of a computational procedure is largely determined by the particular characteristics of the computer which is used to perform the computation. However, there are other factors which must be considered, and some of these are shown in Fig. 1.

Fig. 1 is a block diagram which illustrates the basic steps involved in the solution of the boundary-value problem on automatic computing equipment. Admittedly, this diagram is quite general, since it is applicable to both analogue and digital computing equipment. Correspondingly, the amount of work which must be performed to complete each step in this diagram will vary with the type of computer employed in the computation.

Specifically, if we assume that a complete mathematical statement of the problem is available, the first step is to perform the necessary numerical analysis of the problem. In the case of the digital computer, this phase is concerned with the replacement of the given differential equation by some type of approximation formula, for example, by a suitable set of difference equations, etc. This process may be referred to as the arithmetization of the problem. At this stage a careful study of the probable excursion of the variables must be made to insure that adequate counter capacities are provided. The possibility that a variable may exceed the available capacity of a particular unit on the computer, must be considered and suitable scale factors must be employed.

The second step is concerned with the preparation of the program which is needed in instructing the computer precisely what to do. The actual amount of detailed information which is needed to instruct a computer will vary with the particular machine. A computing machine does not think, and it will exercise only that judgment which has been specifically conveyed to it in the form of instructions (either by virtue of its particular physical design and construction, or by means of programmed instructions). Therefore, the preparation of an accurate program for a digital computer may be an extremely time-consuming task on certain machines. This point is empha-

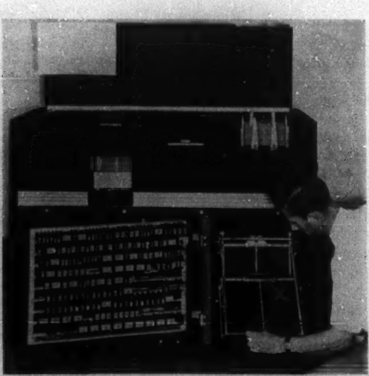


Fig. 2 (left). 602A Electromechanical Calculating Punch

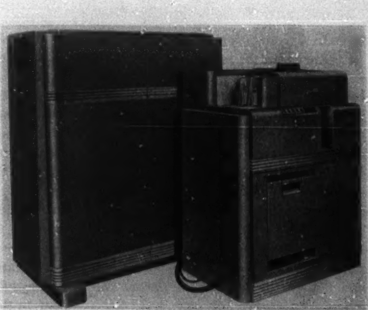


Fig. 3 (right). 604 IBM Electronic Calculating Punch

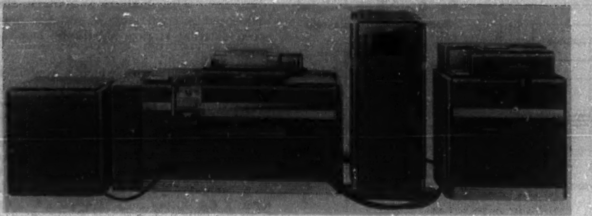


Fig. 4. Card-Programmed Calculator

sized not to detract from the capabilities of such devices; but rather to avoid creating an intoxicating misconception of the capabilities of computers.

The third step on the block diagram is concerned with the insertion of instructions into the computer, that is, actually setting-up the machine to do a particular task. Many different media are used for this purpose; punched paper tape, wired plugboards, punched International Business Machines (IBM) cards, flexowriter tapes, magnetic tapes, etc. In this operation we are concerned with overcoming an effective "language barrier" which exists between the computer and the outside world. That is to say, the computer understands a language which is quite different from that generally spoken. Hence some means of interpretation is needed to overcome the existing language barrier. Again, the actual moment of real time required to effect this operation will vary considerably, depending upon the actual computer under consideration.

The fourth step in the mechanized process is the actual machine solution. Here again, the actual real time expended for a particular example will vary greatly with different computers. Comparative solution times are presented in a later section.

The fifth step is concerned with the checking of the results obtained from the computer. This may consist of an inspection of graphical or tabulated data and an evaluation of the physical significance of the end result. Obviously, the use of a previously-prepared test solution to provide a check of the overall computational program is highly recommended. In the case of the boundary-value problem solved by trial-and-error techniques, a check must be made to determine the agreement between the terminal conditions obtained and the specified second boundary conditions.

The final step in the process is the feedback loop which is always required in an iterative-type solution. In the case of the boundary-value problem, the feedback loop is the operation by means of which the new trial eigenvalue is selected. This step may be a manual operation in which case all available human resources, such as previous experience, judgment, intuition, ingenuity, etc., may be used to an advantage. However, in the machine feedback loop the decision must be based on a previously-determined set of rules. The prime advantage of machine feedback is the possible utilization of the machine's high computing speed.

It is apparent from the preceding introductory remarks that a detailed consideration of each of the individual blocks of Fig. 1 is required. The importance of a well-balanced time distribution among the various steps in the computational process can not be overemphasized!

Brief Physical Description of the Computers

The following photographs are presented in an attempt to provide a concept of the physical dimensions of some of the computers employed in this study. Fig. 2 is a photograph of the *602A* Electromechanical Calculator. The principal instructions for this machine are conveyed to it by means of wires on the control panel, that is, manually inserted wires serve to instruct the calculator to perform the various arithmetic operations of addition, subtraction, multiplication, and division. Twelve separate program steps are available for inserting time-sequenced operations. The *602A* is basically an 8-digit fixed-decimal calculator, and the speed of operation varies between 6 and 50 cards per minute.

Fig. 3 illustrates the physical dimensions of the *604* Electronic Calculator. The calculator operates with 5 decimal-digit numbers at a basic repetition frequency of 50,000 cycles per second. This clock frequency is sufficiently high to permit the execution of 60 different program steps for each punched card passing through the calculator. The machine performs all calculations at a constant speed of 100 cards per minute. In the solution of the boundary-value problem the *604* Calculator is wired to perform 12 decimal-digit arithmetic using a fixed-decimal number system.

Fig. 4 shows the four separate machine units which are interconnected to form the Card-Programmed Calculator (CPC):

1. *412* Accounting Machine;
2. *605* Electronic Calculation;
3. *527* Punch;
4. *941* Storage Unit.

The principal distinction between the CPC and the previously-described *602A* and *604* Calculators is that the CPC receives its instructions in the form of punched cards. Hence, it is no longer necessary to manually wire a control panel for each new computation. One need prepare only the corresponding instruction cards. In this study a so-called general purpose control board was used in the computation. This control panel permits 8 decimal-digit arithmetic using a floating-decimal number system.

Fig. 5 shows the MIT Rockefeller Dif-

Fig. 5. MIT Rockefeller Differential Analyzer

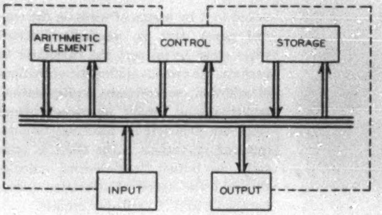


Fig. 6. Typical digital computer block diagram

ferential Analyzer (RDA). This is an analogue machine which is unique because it is the only differential analyzer in existence which receives its instructions on a punched paper tape. It is also the most accurate wheel-and-disc analyzer ever constructed. The machine is designed to provide a maximum accuracy of 5 decimal digits. Specific comments regarding the overall accuracy attained in the solution of the boundary-value problem is given in a later section.

The Functional Characteristics of Computers

During the past decade there has been a tremendous development in the field of digital computation. A host of digital computers have been developed which possess a wide variety of basic characteristics and associated capabilities. Much has been written about these computers, and, therefore, no effort will be made here to repeat what has been said. An attempt will be made, however, to illustrate the relative advantages and limitations of the various computers employed in this investigation.

A digital computer is a device which performs the four fundamental arithmetic operations and makes a few simple decisions upon discrete numerical data. The distinction between the digital computer which employs discontinuous information, and the analogue computer which employs continuous information should be noted. Fig. 6 is a block diagram of the basic computer elements of a digital computer. The diagram illustrates the basic computer elements as separate and distinct units. Actually, this is an oversimplification of the facts. Very often no clear distinction exists between some of these functional blocks; however, as presented, they do provide a basis for a discussion of the relative merits of the different computers.

The over-all effectiveness of digital computers for computational purposes is dependent on the following characteristics:

1. The type of instructions required by the computer. This determines the amount of time required to program the computer.

2. The actual form in which the instructions and input data must be presented to the computer; for example, binary, octal, decimal, etc. This influences the time required to code a problem for the computer.

3. The rate at which the computer accepts data into its internal storage. That is, the speed of the input device affects the time required to set up the computer.

4. The speed with which the arithmetic element performs the different arithmetic operations. This determines the actual internal machine solution time of a problem.

5. The internal high-speed storage capacity of the computer. This characteristic dictates the complexity of the problem which may be conveniently solved on the computer. The distinction between the high-speed internal storage and the low-speed external storage should be emphasized.

6. The accuracy with which a computer performs its operations. This is very important and is largely determined by the actual digital capacity of the registers or counters used in the arithmetic operations.

7. The speed of the output unit. Naturally this strongly influences the capabilities of a computer. Furthermore, the actual form of the output data is of considerable importance.

In addition to the characteristics just mentioned, a true determination of the figure-of-merit of a digital computer must include certain economic considerations such as the initial investment cost, the maintenance and operating costs.

It is evident that a well-designed computer will possess a set of characteristics which is based on a judicious selection of the desirable qualities just mentioned. Admittedly, the ideal computer has yet to be designed; however, each of the succeeding models illustrate remarkable improvements. Since the computers used in this investigation are far from ideal, a just evaluation of their capabilities must include a fair evaluation of the over-all characteristics.

Fig. 7 illustrates the functional characteristics of the *604* Electronic Calculator. The numerical digits indicate the decimal digit capacity of each internal unit. One of the most attractive features of this machine is an 8-decimal-digit channel which permits the internal transfer of information among the various units in a single program step. This provision of a common channel for the data transmission eliminates the need for manually providing a data transmission path of the type required on the *602A* Calculator.

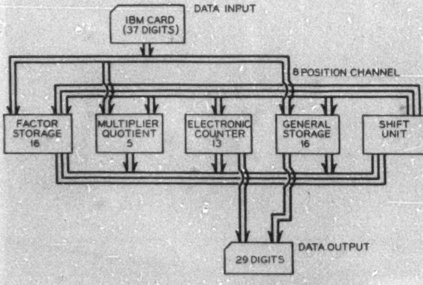


Fig. 7. Block diagram of the 604 Electronic Calculator

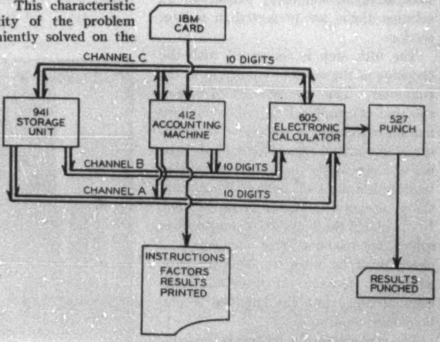


Fig. 8. Block diagram of Card-Programmed Calculator

Furthermore, the existence of undesirable back circuits is thereby precluded, and as a result, the wiring is enormously simplified and the machine set-up time is correspondingly reduced.

Fig. 8 illustrates the functional character of the individual units contained in the CPC. The punched card input to the CPC contains two types of cards, the instruction cards and the data cards. These cards are accepted by the *412* Accounting Machine at a rate 100 or 150 cards per minute. The *412* machine interprets the instruction cards and acts as a control unit for the CPC. If desired, a continuous printed record of the instructions, the factors involved, and the results of the operation is available as the computation progresses. Needless to say, a printed record of this type is extremely valuable to check the results of a computation and also to perform error-diagnostic studies. In this manner, information both before and after the computation is printed for inspection, should the need arise.

The *605* Electronic Calculator performs all of the electronic arithmetic operations, and it may therefore be referred to as the arithmetic element of the CPC. If desired, the results of a computation may be sent to the *527* unit, which punches the answer on IBM punched cards.

Solution of the Boundary-Value Problem

As an illustration of the manner in which a boundary-value problem may be handled on the CPC, consider the solution of the following equation

$$\frac{d^2p}{dx^2}+\left\{2E+\frac{2Z_p}{x}-\frac{l(l+1)}{x^2}\right\}P=0 \tag{1}$$

subject to the boundary conditions

$$P(x=0)=0 \text{ and } P(x=\infty)=0 \tag{2}$$

The physical interpretation of this problem is of no particular concern as far as programming of automatic computing equipment is concerned; hence, no further discussion will be given.

However, if the following definitions are employed

$$G(x)=\left[\frac{l(l+1)}{x^2}-\frac{2Z_p}{x}\right] \tag{3}$$

and

$$\lambda=-2E, \tag{4}$$

then equation 1 may be written as

$$\frac{d^2p}{dx^2}=(\lambda+G)P \tag{5}$$

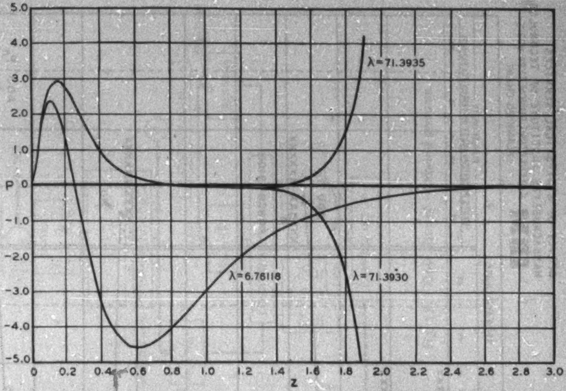


Fig. 9. Solution of Schrodinger's equation using Taylor's series formula

Card-Programmed Calculator

Solution	Energy Level	Eigenvalue, λ
1	2P	71.3935
2	2P	71.3930
3	3P	6.76118

The next step is to replace the second-order derivative by a suitable difference formula—a formula designed to minimize the truncation error (the first term ignored in a power-series expansion of a function).

A new variable

$$y_n=p_n-\frac{h^2}{12}\left(\frac{d^2p}{dx^2}\right)_n \tag{6}$$

is now chosen, which is designed to eliminate the fourth derivative in a Taylor-series expansion of the function y_n. Using the familiar second-order formula

$$\Delta_n{}''y=y_{n+1}-2y_n+y_{n-1} \tag{7}$$

and a Taylor-series expansion for the functions y_n, we obtain

$$\Delta_n{}''y=h^2\left(\frac{d^2p}{dx^2}\right)_n-\frac{h^6}{240}\left(\frac{d^6p}{dx^6}\right)_n+\ldots \tag{8}$$

If only the first term in equation 8 is used to approximate the second derivative, the truncation error thus incurred is given by

$$T\simeq\frac{h^6}{240}\left(\frac{d^6p}{dx^6}\right)_n \tag{9}$$

If this approximation is applied to equation 5, equations 10 are obtained

$$P_n=\left[\frac{1}{1-\frac{h^2}{12}(\lambda+G_n)}\right]Y_n$$

$$\Delta_n{}''Y=h^2(\lambda+G_n)P_n$$

$$\underset{n+1/2}{\Delta}{}'Y=\Delta_n{}''Y+\underset{n-1/2}{\Delta}{}'Y \tag{10}$$

$$Y_{n+1}=\underset{n+1/2}{\Delta}{}'Y+Y_n$$

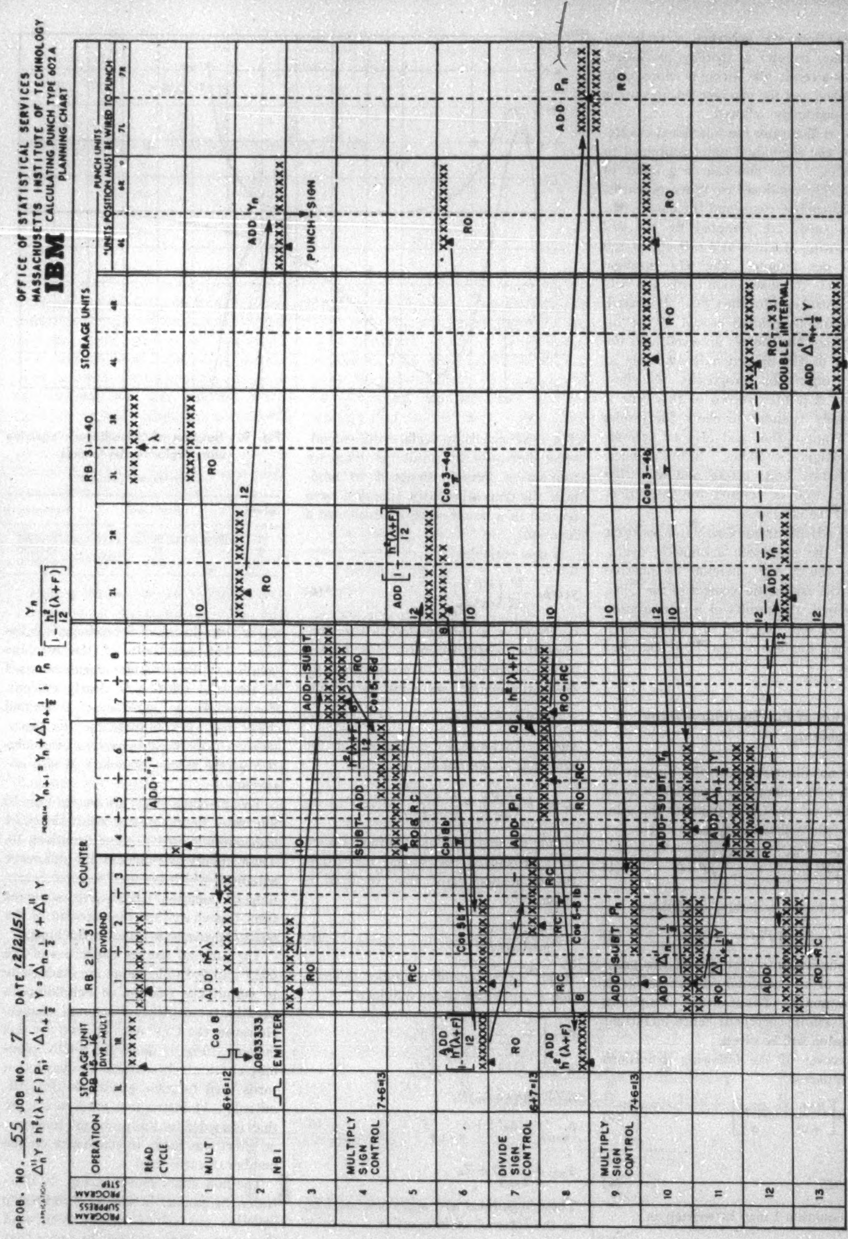
These equations are applicable as long as the interval of integration h remains constant.

Fig. 9 illustrates the graphical character of the solution of Schrödinger's equation. The sensitivity of the machine solution to the particular eigenvalue used in the trial solution is clearly evident. For example, an accuracy of 6 decimal digits may be obtained for the eigenvalue. The critical behavior of the solution at the second boundary is also apparent.

The planning chart shown in Fig. 10 illustrates the manner in which the *602A* Calculator is used to solve equations 10. These equations represent the difference equations which are used to solve Schrödinger's equation. It is quite apparent from this chart that the capacity of the *602A* Calculator is taxed to the limit.

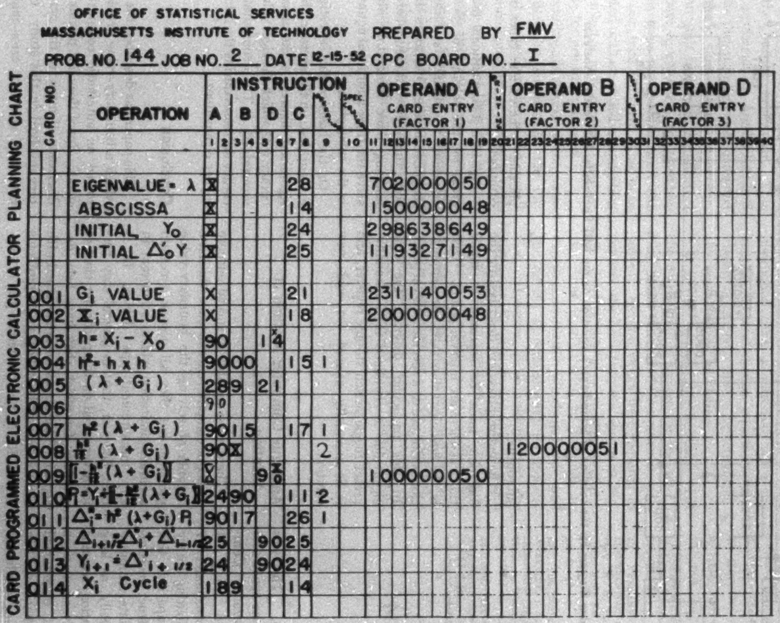
The internal storage capacities of the *602A* and *604* Calculators are inadequate to permit the solution of Schrödinger's equation at a point of interval change. However, the CPC may be used for this purpose. Fig. 11 illustrates a CPC planning chart which contains 14 instruction cards used to solve equations 10. Admittedly, 14 instruction cards are more than is needed for this problem; however, no effort was made to economize on the number of cards used.

The flow chart shown in Fig. 12 illustrates the manner in which Schrödinger's equation was solved on the Whirlwind Computer. First, it must be noted that both single- and double-register-length



OFFICE OF STATISTICAL SERVICES
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
IBM
CALCULATING PUNCH TYPE 602 A
PLANNING CHART
PROB. NO. 55 JOB NO. 7 DATE 12/21/51
PROGRAM SUPPRESS
PROGRAM STEP
OPERATION
STORAGE UNIT RB 5-16
DIVR-MULT
COUNTER
RB 21-31
DIVIDEND
RB 31-40
STORAGE UNITS
PUNCH UNITS
READ CYCLE
MULT
NBI
MULTIPLY SIGN CONTROL
DIVIDE SIGN CONTROL
EMITTER
.0833333
ADD
SUBT-ADD
ADD-SUBT
RO
RC
PUNCH-SIGN
ROT-X31 DOUBLE INTERVAL
PREPARED BY

Fig. 10. Solution of Schrodinger's equation using Taylor's series formula



OFFICE OF STATISTICAL SERVICES
MASSACHUSETTS INSTITUTE OF TECHNOLOGY PREPARED BY FMV
PROB. NO. 144 JOB NO. 2 DATE 12-15-52 CPC BOARD NO. I

CARD PROGRAMMED ELECTRONIC CALCULATOR PLANNING CHART

CARD NO.	OPERATION	INSTRUCTION A (1 2)	INSTRUCTION B (3 4)	INSTRUCTION D (5 6)	INSTRUCTION C (7 8)	9	10	OPERAND A CARD ENTRY (FACTOR 1) (11–19)	OPERAND B CARD ENTRY (FACTOR 2) (21–30)	OPERAND D CARD ENTRY (FACTOR 3) (31–40)
	EIGENVALUE = λ	X			28			702000050		
	ABSCISSA	X			14			150000048		
	INITIAL Y_0	X			24			298638649		
	INITIAL $\Delta'_0 Y$	X			25			119327149		
001	G_i VALUE	X			21			231140053		
002	X_i VALUE	X			18			200000048		
003	$h = X_i - X_0$	90		$1\overset{x}{4}$						
004	$h^2 = h \times h$	90	00		15	1				
005	$(\lambda + G_i)$	28	9	21						
006		90								
007	$h^2(\lambda + G_i)$	90	15		17	1				
008	$\frac{h^2}{12}(\lambda + G_i)$	90	X			2			120000051	
009	$\left[1 - \frac{h^2}{12}(\lambda + G_i)\right]$	X		$9\frac{X}{0}$				100000050		
010	$P_i = Y_i \div \left[1 - \frac{h^2}{12}(\lambda + G_i)\right]$	24	90		11	2				
011	$\Delta''_i = h^2(\lambda + G_i)P_i$	90	17		26	1				
012	$\Delta'_{i+1/2} = \Delta''_i + \Delta'_{i-1/2}$	25		90	25					
013	$Y_{i+1} = \Delta'_{i+1/2}$	24		90	24					
014	X_i Cycle	18	9		14					

Fig. 11. Solution of Schrodinger's equation on the card-programmed calculator

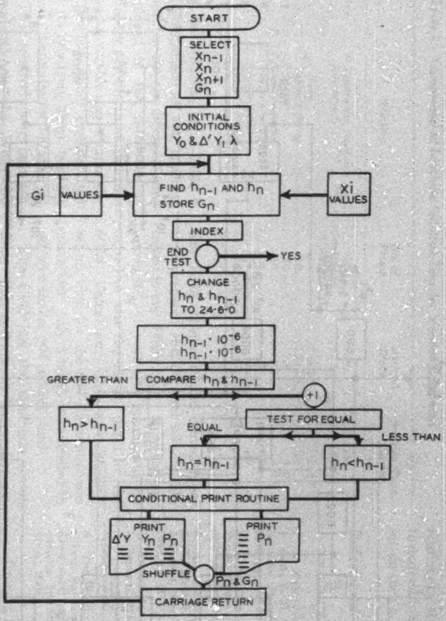

START
SELECT X_{n-1} X_n X_{n+1} G_n
INITIAL CONDITIONS Y_0 & $\Delta' Y_1$ λ
Gi VALUES
FIND h_{n-1} AND h_n STORE G_n
Xi VALUES
INDEX
END TEST
YES
CHANGE h_n & h_{n-1} TO 24-6-0
$h_{n-1} \cdot 10^{-6}$
$h_{n-1} \cdot 10^{-6}$
COMPARE h_n & h_{n-1}
GREATER THAN
+1
$h_n > h_{n-1}$
TEST FOR EQUAL
EQUAL
LESS THAN
$h_n = h_{n-1}$
$h_n < h_{n-1}$
CONDITIONAL PRINT ROUTINE
PRINT $\Delta' Y$ Y_n P_n
PRINT P_n
SHUFFLE
P_n & G_n
CARRIAGE RETURN

Fig. 12. Flow chart of whirlwind program

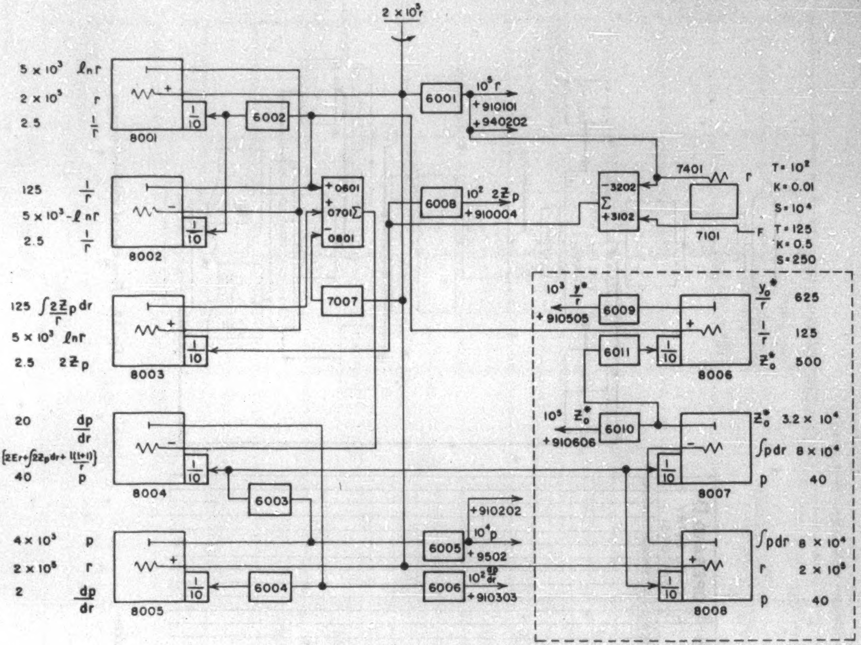

2 × 10³r
5 × 10³ ln r
2 × 10³ r
2.5 1/r
8001
6001
10⁵r
+910101
+940202
6002
125 1/r
5 × 10³ − ln r
2.5 1/r
8002
+0601
+0701Σ
−0801
6008
10² 2Z p
+910004
−3202
Σ
+3102
7401
r
7101
F
T = 10²
K = 0.01
S = 10⁴
T = 125
K = 0.5
S = 250
125 ∫2Z p/r dr
5 × 10³ ln r
2.5 2Z p
8003
7007
10³ y*/r
+910505
6009
6011
y₀*/r 625
1/r 125
Z₀* 500
8006
20 dp/dr
30 − {2Er + ∫2Z p dr + l(l+1)/r}
40 p
8004
6003
10⁵ Z₀*
+910606
6010
Z₀* 3.2 × 10⁴
∫pdr 8 × 10⁴
p 40
8007
+910202
10⁴p
+9502
4 × 10³ p
2 × 10⁵ r
2 dp/dr
8005
6005
6004
6006
10² dp/dr
+910303
∫pdr 8 × 10⁴
r 2 × 10⁵
p 40
8008
1/10

Fig. 13. Machine diagram for the solution of Schrodinger's equation on the MIT Rockefeller Differential Analyzer

Given equation

$$\frac{d^2p}{dr^2}+\left[2E+\frac{2Z_p}{r}-\frac{l(l+1)}{r^2}\right]p=0$$

Machine equations

$$\frac{dp}{dr}=-\int p\,d\left[2Er+\int\frac{2Z_p}{r}dr+\frac{l(l+1)}{r}\right]$$

$$Z_0{}^*=\int p^2dr$$

$$\frac{y_0{}^*}{r}=\int Z_0{}^*d\left(\frac{1}{r}\right)$$

$$2Z_p=Kr+F$$

numbers were employed. This was necessary to conserve storage space and still provide the desired accuracy. Computations were performed using 24 binary-digit numbers (approximately 7 decimal digits) and a floating-binary number system. The comparison test is used to automatically instruct the computer to select an appropriate computational subroutine. Specifically, as long as the interval of integration remains constant, the computer solves equations 10. However, at a point of interval change a different set of equations, 11 in number, must be solved, using a different subroutine.

The entire procedure described in the foregoing must be repeated as many times as there are interval points. That is, the flow chart illustrates the computation performed at a specific interval point. In this particular diagram the trial eigenvalue is contained in a manually-preset toggle-switch storage register, and it is manually altered for each new solution.

Analogue Computer Method of Solution

There are a number of special-purpose analogue computers in existence which are capable of solving the boundary-value problem, but they all possess a certain accuracy limitation which is inherent in any measurement-type device. It is true, however, that in general they perform very well on the particular class of problems for which they have been designed. This discussion is concerned with the solution of boundary-value problems on the RDA, but the techniques are applicable to the various types of analogue equipment.

Fig. 13 is a conventional machine diagram for the solution of Schrödinger's equation. It should be observed that only five of the integrators are used to solve the basic equation, whereas the remaining three integrators are used to solve the three additional equations listed in the subcaption.

On the basis of this investigation it is evident that, on the average, three or four significant figures is the maximum accuracy with which the eigenvalues may be ascertained. The computational accuracy decreases during the solution of the higher eigenvalues.

Conclusion

This paper has described the manner in which a boundary-value problem can be solved on automatic computing equipment. Space does not permit a detailed discussion of many of the pertinent factors involved in this study, that is, accuracy of solution, error-analysis tech-

Table I. Comparison of Machine Solution Time for Various Computers

Computer	Solution	No. of Points	Decimal Digits	Solution Time, Minutes
Desk Calculator	Quadrature Formulae	192	6	1440
RDA	Analogue	Continuous 3 Ranges	5	60
602A-I	3-point	150	7 to 12	12.5
602-III	Taylor's Series	150	7 to 12	25
604	3-point	150	10	3.0
604	Taylor's Series	150	10	3.0
CPC	Taylor's Series	150	8	20
CPC	Runge-Kutta	150	8	40
WW1	Taylor's Series	150	7	19

niques, cost per solution, etc. Table I contains a comparison of the machine solution time for the six different computers employed in the study. It is apparent that the *604* Electronic Calculator is the most favorable computer for the particular problem under consideration because of its relatively short machine solution time (3 minutes for 150 interval points). The relatively-long solution time encountered on the Whirlwind Computer was predicated by the limitation in the output typewriter. This work was performed in April 1952, and at that time the output element consisted of a relatively slow output typewriter. This output unit increased the effective solution time enormously; for example, 19 minutes were required to print the results of one trial solution. This is to be contrasted with an actual computation time of 2 minutes, indicating a 17 minute loss of time for printing purposes.

The preceding discussion of computational philosophy emphasized the importance of a consideration of the overall computational problem of the type indicated in Fig. 1. It is therefore imperative that the problem preparation time be considered. Generally speaking, the time required to set up a digital computer is greater than that required for an analogue computer. Specifically, the necessary RDA schematics and control tapes were prepared and the first solution obtained in one 8-hour day. However, approximately 1 week (40 hours) was required to wire the control panel for the *602A* Calculator. The corresponding set-up time for the *604* Calculator was much less—only 8 hours were needed to set up a similar *604* control panel.

The CPC has the minimum set-up time. For example, 3 hours were required to prepare the planning charts and instruction cards required to solve the general interval-change equations previously described. Hence an actual solution was obtained in 4 hours.

It is with considerable hesitancy that the corresponding figures for the set-up time of the Whirlwind Computer are presented. Several weeks were required to prepare the final program shown in Fig. 12. Part of this long problem preparation time is attributed to personal inexperience with the Whirlwind code, and the details associated with its application. However, it is evident that the coding of the digital computer in which actual "bits" of numbers are to be manipulated is necessarily more exacting than a code in which arithmetic and trigonometric operations are simply coded with a single operational code. Clearly, the recent development of the Library of Subroutines by the Whirlwind staff will serve to materially decrease the time required to set-up such a problem.

References

ANALOGUE COMPUTING EQUIPMENT

1. A NEW TYPE OF DIFFERENTIAL ANALYZER, V. Bush, S. H. Caldwell. *Journal*, The Franklin Institute, Philadelphia, Pa., vol. 240, no. 4, Oct. 1945, pp. 255–326.

2. A GENERALIZED ANALOGUE COMPUTER FOR FLIGHT SIMULATION, Albert C. Hall. *AIEE Transactions*, vol. 69, pt. I, 1950, 308–20.

3. AN ELECTRONIC DIFFERENTIAL ANALYZER, A. B. Macnee. *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 37, no. 11, Nov. 1949, pp. 1315–24.

4. LINEAR ALGEBRAIC SYSTEMS AND THE REAC, L. Gephart. Mathematical Tables and Other Aids for Computation, National Resources Council, Washington, D. C., vol. 6, July 1952, pp. 190–203.

5. THE SOLUTION OF A CLASS OF LINEAR OPERATIONAL EQUATIONS BY METHODS OF SUCCESSIVE APPROXIMATION, J. M. Ham. Sc.D. Thesis, Massachusetts Institute of Technology, Cambridge, Mass., 1952.

DIGITAL COMPUTING EQUIPMENT

6. IBM CALCULATORS. *Proceedings*, IBM Computation Seminar, Aug. 1948.

7. Ibid., Nov. 1949.

8. Ibid., Dec. 1949.

9. Ibid., Sept. 1950.

10. Ibid., Aug. 1951.

11. THE WHIRLWIND I COMPUTER, R. R. Everett. *Electrical Engineering*, vol. 71, Aug. 1952, pp. 681–86.

12. FUNDAMENTAL CONCEPTS OF THE DIGITAL DIFFERENTIAL ANALYZER METHOD OF COMPUTATION, R. E. Sprague. Mathematical Tables and Other Aids for Computation, National Resources Council, Washington, D. C., vol. 6, Jan. 1952, pp. 41–49.

13. THE SERIAL-MEMORY DIGITAL DIFFERENTIAL ANALYZER, J. F. Donan. Mathematical Tables and Other Aids for Computation, National Resources Council, Washington, D. C., vol. 6, April 1952, pp. 102–12.

14. SOLUTION OF BOUNDARY VALUE PROBLEMS ON AUTOMATIC COMPUTING EQUIPMENT, F. M. Verzuh. Sc.D Thesis, Massachusetts Institute of Technology, Cambridge, Mass., 1952.

◆

No Discussion

Nondestructive Sensing of Magnetic Cores

DUDLEY A. BUCK
NONMEMBER AIEE

WERNER I. FRANK
NONMEMBER AIEE

Synopsis: A nondestructive method for sensing the information state of magnetic-memory cores is under development. This new method involves the use of a quadrature magnetic field which, when pulsed, produces an output voltage. The polarity of the output voltage contains the information: a positive output pulse represents a ONE, and a negative pulse a ZERO. Reading in this manner can be repeated indefinitely without destroying the information held by the core. The method has proven to be exceedingly fast, giving output pulses of 0.2 microsecond in duration, and because of the rapid flux change when sensing, the output voltage has been observed in excess of 1 volt per turn.

MAGNETIC cores with rectangular hysteresis loops are of interest to the field of digital computers because of their ability to hold, or "remember," the direction in which they have been magnetized. In digital computer terminology, a magnetic core may be said to contain a binary ZERO if its remanent flux is positive or a ONE if its remanent flux is negative. A group of magnetic cores can therefore hold a pattern of ZEROS and ONES which make up a binary number.

How Information is Stored in and Read out of a Magnetic Memory Core

Fig. 1 shows a single magnetic memory core together with a somewhat idealized plot of its rectangular hysteresis loop, i.e., a graph of the amount of magnetic flux (Φ) present in the core as a function of the magnetizing force (NI) impressed on that core.

To write a ZERO in the core, a pulse of current is made to flow in the magnetizing winding in a positive direction, setting up magnetic flux in a clockwise direction. This corresponds to a movement to the right (positive direction) on the Φ versus NI graph. It can be seen that if we move far enough to the right, following the curve, we will eventually end up on the upper horizontal slope of the hysteresis loop. We will have saturated the core, and regardless of how much the magnetizing force is increased, the flux in the core remains the same. When the magnetizing force is removed corresponding to a displacement to the left in the Φ-NI plane, we move along the upper horizontal slope until we come to rest at its intersection with the center line (no magnetizing force) marked "0." The graph indicates that almost the entire saturation flux is retained even after the magnetizing force has been removed. The core now holds a ZERO.

Conversely, in order to write a ONE, negative current sets up negative flux, i.e., counterclockwise. In terms of the Φ-NI plane this means that the core is moving to the left along the curve until it hits the bottom horizontal slope. When the magnetizing current is stopped, the core comes to rest at the intersection of the bottom horizontal and the vertical centerline. The core then holds a ONE.

The read process makes use of the fact that when a magnetic flux change occurs inside a coil, a voltage is generated in that coil. An output winding is placed on the memory core (bottom of sketch in Fig. 1), and a sensing pulse is applied to the input winding. The sensing pulse, also called "read" pulse, indeed, is nothing but a "write-ZERO" pulse. When it is applied, two possibilities exist: either the core holds a ONE or it holds a ZERO.

If the core holds a ONE, it starts out at the bottom, center, of the hysteresis loop marked "1" and begins to move to the right. The extent of this motion is indicated by the arrow marked "sensing pulse." As the core gets around the knee of this curve, a large change of flux occurs while we move from the bottom to the top of the rectangular hysteresis loop. This large change of flux generates a large voltage in the output winding, thus indicating that the core had held a ONE.

If the core holds a ZERO, (i.e., rests at the top, center, of the hysteresis loop) and a sensing pulse is applied, all that happens is that the core moves out to the right on the graph along an essentially horizontal line. Thus very little flux change occurs; therefore no voltage, or only a very small voltage, appears across the output winding, thus indicating that the core had held a ZERO.

A somewhat simpler way of looking at this read process is to imagine that it consists in magnetizing or "switching" the core into the ZERO position. If the core was initially in the ONE state, the flux has to be reversed, and therefore a change of flux occurs which generates an output voltage. If the core was already in the ZERO position, no change of flux is needed and no output signal is generated.

As shown in the foregoing, this system of reading destroys the stored information: after having been read, a core will always find itself in the ZERO, or "cleared" state regardless of whether it contained a ONE or a ZERO just prior to reading. If information is to be retained for further use, it must be rewritten into the core. The challenge, quite naturally, arose to devise a system of extracting information which would not destroy that information in the process. This paper describes one such nondestructive read system.

A Nondestructive Read System

This particular nondestructive read system is based on the tendency of materials with rectangular hysteresis loops to maintain their remanent flux in certain easy directions of magnetization. These materials behave in many ways like a single ferromagnetic crystal, or cube, with easy directions of magnetization along its edges. In the absence of an external field, the residual magnetization vector of such a cube will lie in one of these easy directions (Fig. 2, left). With this diagram reduced to a 2-dimensional picture (Fig. 2, right), a hard direction of magnetization can be imagined between each of the easy directions, as shown. If a magnetic field were applied at right angles to the remanent-flux vector, the vector would tend to align itself with the field and would thus be rotated away from its

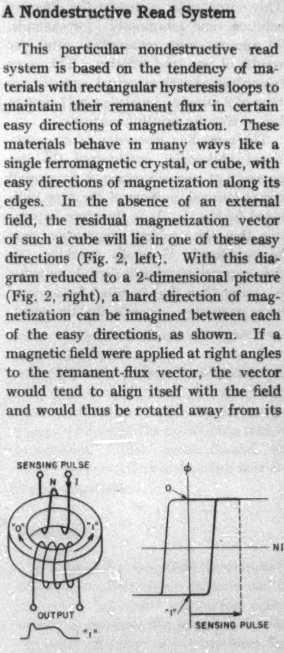


Fig. 1. Destructive sensing

Paper 53-409, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Kansas City, Mo., November 2–6, 1953. Manuscript submitted June 2, 1953; made available for printing October 2, 1953.

DUDLEY A. BUCK and WERNER I. FRANK are with the Massachusetts Institute of Technology, Cambridge, Mass.

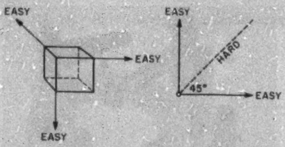


Fig. 2. Easy directions of magnetization

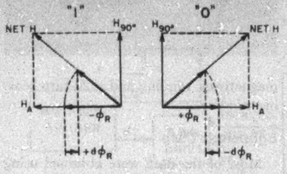


Fig. 3. Vector representation of nondestructive sensing

preferred direction toward a hard direction of magnetization. If this quadrature field were not strong enough to rotate the remanent-flux vector past the hard direction, this vector would snap back to its original easy direction as soon as the quadrature field was removed. (For easier identification, the usual magnetizing field on a core will henceforth also be referred to as the α field, while the quadrature field will also be called a β field.)

This same principle may be used in conjunction with larger masses of rectangular-hysteresis-loop materials. These materials may be thought of as made up of individual crystals all oriented in such a manner that one of their easy directions of magnetization lies in the direction in which the magnetizing (α) field is to be applied. In the case of a ring-shaped core, this means that one of the easy directions would be the tangential.

Fig. 3 illustrates what happens when a magnetic field is applied at right angle to the remanent flux of a memory core. The remanent flux (Φ_R) originally lies along the α direction (in line with $\mathbf{H}_A$). If the core contains a ONE, this will be in the negative direction ($-\Phi_R$); if it contains a ZERO, in the positive direction ($+\Phi_R$). When the quadrature magnetic field ($H_{90\text{ degrees}}$) is applied at right angle to $\mathbf{H}_A$, the two will combine, vectorially, to form a resultant field (net $\mathbf{H}$). The remanent flux Φ_R will then line up with this combined field; i.e., it is rotated away from its easy direction of magnetization (along the dotted arc in Fig. 3). The magnitude of the flux, however, will remain essentially unchanged, since in a rectangular-hysteresis-loop material the remanent flux is practically the same as the full saturation flux. Therefore, the output winding on that core effectively will see the flux reduced from Φ_R to $\Phi_R \cos\theta$ (where θ is the angle of vector rotation), and a voltage is induced in the output winding. In the case of a zero with the remanent flux in the positive direction, the winding will see a negative change of flux ($-d\Phi_R$) as the β field is applied, and a positive change of flux as the β field is removed and the remanent flux is allowed to snap back into its preferred direction. Conversely, in the case of a ONE with the remanent flux in the negative direction, a positive voltage will be induced upon application of the β field, and a negative voltage upon removal of the β field. Thus, merely by observation of the polarity of the voltage induced in the winding of the memory core upon application of a β field, the state of magnetization of this core may be ascertained. In other words, the memory core may be read without destroying the information it contains.

It might be well to note here that this sensing operation does not switch the magnetic flux in the core, and that therefore it does not involve the destruction and formation of domain boundaries. This means that very little hysteresis loss is incurred and that the system is, potentially, very fast.

Experimental Setups and Techniques

As soon as it was attempted to try the proposed nondestructive read scheme, the question arose as to how to generate the β field. Two possibilities were envisaged: an externally generated β field, and an internally generated β field.

The Externally Generated β Field

A magnetic path, made of iron or some other material without a rectangular hysteresis loop, is so constructed that it contains a gap into which a memory core may be inserted (Fig. 4). Through a winding on the iron path, a short current pulse is driven which induces magnetic flux at right angles to the remanent flux of the memory core. The remament-flux vector is rotated for as long as the quadrature-field current flows, and the sensing winding (or output winding) sees the change of flux in the tangential direction. This change of flux, ideally, occurs just as rapidly as the current in the β-field winding can build up. Now, the voltage induced in the sensing winding is proportional to the time rate of change in flux which this winding sees; so that, although the flux change is relatively small, the voltage output might be quite large if the change can be made to take place in a very short time.

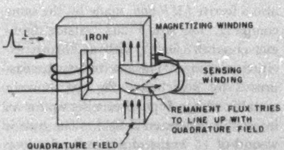


Fig. 4. Externally generated β field

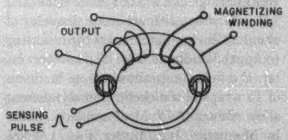


Fig. 5. Hollow toroid geometry

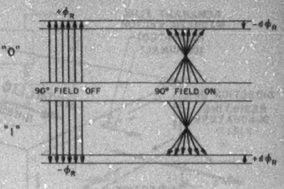


Fig. 6. Vector representation of hollow toroid sensing

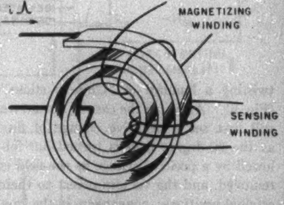


Fig. 7. Metallic-ribbon-type memory core connected to generate β field internally

The Internally Generated β Field

Consider a hollow, toroidal magnetic-memory core (Fig. 5) with a conductor placed circumferentially inside the hollow. The magnetizing winding occupies its usual place, and, after current has been allowed to flow through it, residual flux will exist along the large circumference. When a current pulse is sent through the conductor inside this hollow toroid, a field will temporarily exist around the conductor, deflecting the circumferential remanent flux in either a clockwise or counterclockwise direction. The effect may be compared to that obtained by

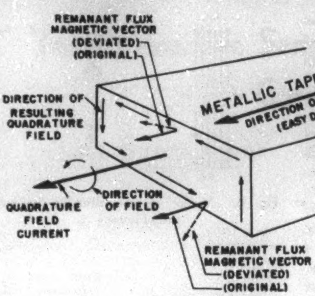

REMANANT FLUX MAGNETIC VECTOR
(DEVIATED)
(ORIGINAL)
DIRECTION OF RESULTING QUADRATURE FIELD
METALLIC TAPE
DIRECTION OF ROLLING (EASY DIRECTION OF MAGNETIZATION)
DIRECTION OF FIELD
QUADRATURE FIELD CURRENT
REMANANT FLUX MAGNETIC VECTOR
(DEVIATED)
(ORIGINAL)

Fig. 8 (left). Internally generated quadrature field

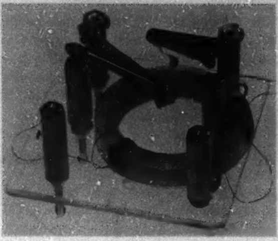


Fig. 10 (right). 1-mil Deltamax used with internal β field

twisting a handful of "Chinese sticks" (Fig. 6). Again, in the case of a ZERO we will first see a negative change of flux as the 90-degree field is turned on, followed by a positive change as the field is removed, and the vectors revert to their original positions; conversely, in the case of a ONE we will see a positive change followed by a negative one.

K. H. Olsen suggested, and first tried, a different way of generating internal β fields in ribbon-wound cores. With leads attached directly to the two ends of the metallic tape (Fig. 7), the core itself forms the path for the quadrature-field current.

In Fig. 8 is shown a perpendicular cross section of the metallic tape. The large arrow protruding from the face of the section represents a quadrature-field-current vector surrounded by its associated magnetic field (counterclockwise). Since the entire tape conducts current, the total resultant quadrature field will be strongest around the outside edges of the section (represented by the six thin arrows in the plane of the section). The remanent-flux vectors perpendicular to the plane of the section are deviated by this field as shown; again a sensing winding will detect the change of flux.

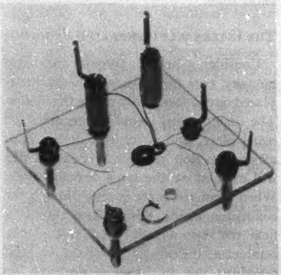


Fig. 9. Arrangement for an externally generated β field

Other Quadrature Fields

The possibility was also envisaged of using a third field at right angles to both the initial remanent flux (α field) and the 90-degree field (β field). This might be called the γ field and might be useful for some sort of coincident-field selection scheme, in which, for instance, the β field is pulsed along the rows and the γ field along the columns of a matrix. A ONE, say, would be indicated by coincidence of pulses on two output windings, one sensitive to β excitation and the other to γ excitation.

Typical Setups

Fig. 9 shows a typical arrangement for an externally generated β field: The memory core is a rectangular-loop ferrite (*MF-1118*, made by General Ceramics and Steatite Corp., Keasbey, N. J., die size *F-259*) with a 1-turn magnetizing winding and a 20-turn sensing (output) winding. The quadrature field is developed across a slot cut in a larger core, also a ferrite (*MF-666*, made by the same company, die size *F-262*, outside diameter about 3/8 inch), with a 20-turn exciting winding. The Lucite board measures 3 by 3 inches. Fig. 10 shows a metallic-ribbon type of core set up for an internally generated β field. The core is wound of 15 wraps of a rectangular-loop nickel-iron alloy (Deltamax, made by Allegheny - Ludlum Steel Corporation, Brackenridge, Pa.) 0.001 inch thick and 1/4 inch wide, with an inside diameter of about 2 inches, and a 20-turn sensing (output) winding. Fig. 11, another internal β-field setup, shows a core made up of 10 wraps of a nickel-iron-molybdenum alloy ribbon (*4-79* Mo-Permalloy, made by Magnetics, Inc., Butler, Pa.), 1/4 mil thick and 1/8 inch wide, with a 2-turn magnetizing winding and a 20-turn sensing winding.

Equipment Used

Most of the data were obtained using the core-testing setup shown in Fig. 12. This arrangement is made up largely of Massachusetts Institute of Technology (MIT) Digital Computer Laboratory standard test-equipment building blocks. It furnishes, for magnetizing and sensing circuits, three pulses, one of maximum amplitude 750 milliamperes ($\mathbf{H}_m$) and two each with a maximum of about 400 milliamperes ($1/2\mathbf{H}_m$), which may be combined to form one pulse of 800 milliamperes.

The no. 3 pulse will always follow a no. 1 pulse (unless, of course, the switches 1 or 2 are opened), after which from 0 to well over 100 no. 2 pulses may be inserted. The length of these pulses may be varied by adjusting the time delay between the input to and the output from the gate and delay unit no. 2 (near the bottom of the block diagram of Fig. 12) from about 0.2 microsecond to well over 100 microseconds. For every particular setting, all three pulses will be of equal width. The repetition rate is varied by changing the free-running frequency of the high-frequency multivibrator in the multivibrator

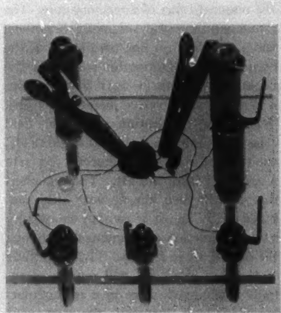
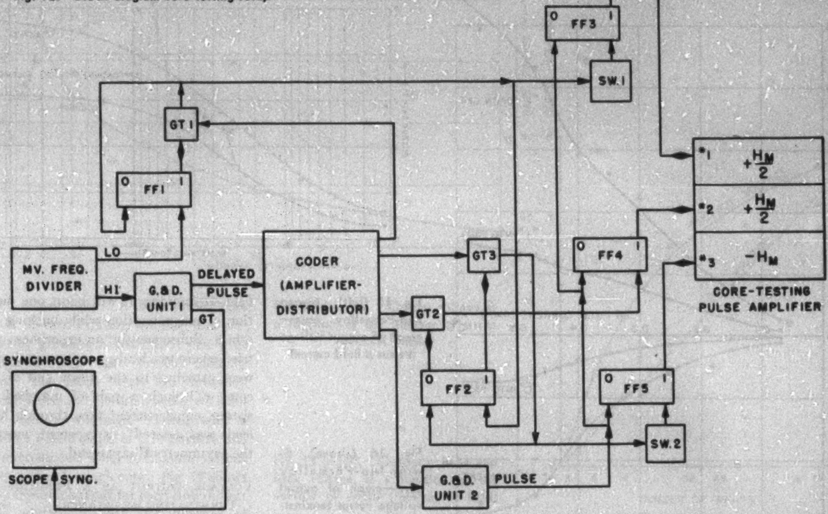


Fig. 11. 1/4-mil Mo-Permalloy with internal β field

Fig. 12. Block diagram core-testing setup



frequency divider from about 2 to about 250 kc.

The amplitude of the current pulses is variable from 0 to about 750 milliamperes. For those tests requiring larger currents, a thyratron pulse generator was used varying from about 0.8 ampere to 3.8 amperes. The number of no. 2 pulses inserted between nos. 3 and 1 is varied by varying the low frequency of the multivibrator frequency divider.

Oscilloscopes

All measurements (except d-c resistance measurements), readings, and photographs were taken on either a type *513D* or a type *514D* Tektronix scope, both having a 1-megohm and 40-micromicrofarad input impedance.

Steady-State Measurements

For steady-state measurements, the core was switched into ZERO and ONE position by means of a battery, a key, a variable resistor, and a reversing switch. Once the information had thus been stored, only one pulse was needed: the quadrature-field readout pulse.

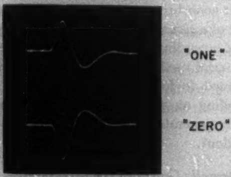


Fig. 13. Typical wave forms in an unterminated output winding

Transient Measurements

For transient measurements, again, the information was stored, initially, by means of the outside battery. Then, with *SW* 1 open, the quadrature-field leads were connected to pulse no. 2 and the magnetizing winding of the core to pulse no. 3. Thus a variable number of readout pulses could be set up, followed by a remagnetizing pulse.

Experimental Results

Typical Wave Forms

Fig. 13 shows typical wave forms (voltage versus time) obtained with an open output winding. The top wave form shows the output of a memory core holding a ONE. It illustrates the positive change in flux (represented by the positive voltage induced in the sensing winding) as the β field is turned on, followed by a negative change as the β field is removed and the remanent flux returns to its easy direction of magnetization. When the information held by the core is changed to a ZERO (Fig. 13, bottom), a pattern of opposite polarity is observed.

Fig. 14 shows the output of a metallic-ribbon type of core into a 20-turn sensing winding terminated in 93 ohms, with both internally and externally generated β field. As may be seen, the loading of the output winding has the effect of slowing down the return of the disturbed remanent-flux vector to its easy direction of magnetization, so that the output signal is essentially confined to a single positive pulse for a ONE and a negative pulse for a ZERO.

There is a striking similarity between

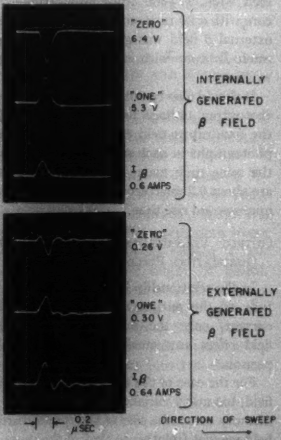
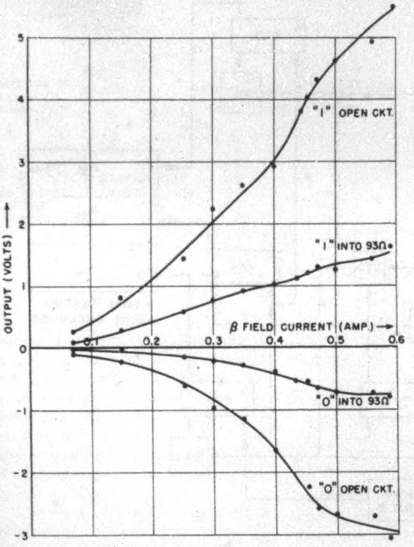


Fig. 14. Typical wave forms in an output winding terminated with 93 ohms



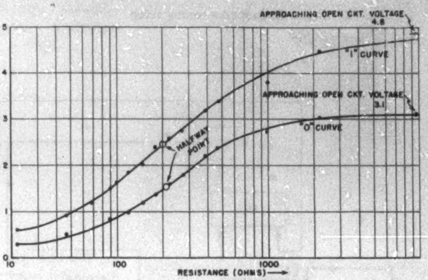


Fig. 15 (left). 6-wrap Mo-Permalloy core—graph of output voltage versus β field current

Fig. 16 (above). 6-wrap Mo-Permalloy core—graph of output voltage versus terminating resistance

results obtained with the two different methods of generating the β field. Except for a scale factor and some oscillation introduced by the external β field, the results may be said to be alike. This is as it should be: the output of the memory core should not depend on the manner in which the quadrature field pulse is generated. Material used is a 40-wrap, 1/8-inch wide, 1/4-mil thick Mo-Permalloy core with a 20-turn output winding. The external β field was generated in a Ferramic-*D* core with a 10-turn exciting winding.

It should also be noted here how closely the output voltage follows the shape of the quadrature-field current pulse. The photographs in each series were taken to the same time scale; all of these pulses are about 0.2 microsecond long with a 0.1-microsecond rise time.

OUTPUT VOLTAGE AS A FUNCTION OF QUADRATURE FIELD CURRENTS

In the relationship between the magnitude of the output voltage and β-field current, only the internally generated field offers any sensible basis for comparison.

For the case of the external quadrature field, too many variables exist outside the memory core; the size of the quadrature field magnet (relative to that of the memory core), its magnetization curve, the amount of residual flux it retains, and the smoothness of the mechanical fit between it and the memory core are just a few of these variables. With the internally generated β field, all these parameters are fixed for each core.

Fig. 15 shows the output into 93 ohms and open-circuit voltages of a 6-wrap Mo-Permalloy core 1/4 mil thick, 1/8 inch wide, pulsed with a current pulse 0.2 microsecond long, the amplitude of which was varied from about 0.07 ampere to over 0.6 ampere. The output winding has 20 turns. The curves are typical of many others, obtained both with externally and internally generated β fields.

There is a long, essentially linear (or slightly concave) rise of output voltage as the β-field current is increased, followed by a flattening out and finally (not shown in this graph) a rapid decrease of output as the nondestructive limit is exceeded. The decrease is probably caused by the quadrature field deflecting the remanent-flux vectors ϕ_n more than 45 degrees from their original position along $\mathbf{H}_A$ (see Fig. 3), so that they then line up along $\mathbf{H}_{90\text{degrees}}$ instead of dropping back to the $\mathbf{H}_A$ direction.

A certain asymmetry is noticeable in these curves; the ONE outputs are consistently larger than the ZEROS. This is caused by the β-field-current lead attached to the inside end of the metallic-ribbon memory core; since the current flows in from one side only, the magnetic field surrounding it will boost one direction of magnetization while bucking the other. Subsequently, an experiment was tried where two leads, one from each side, were attached to the inside end of the core. Through a pair of matched resistors, equal current flow through both leads was assured; as expected, most of the asymmetry disappeared.

VOLTAGE OUTPUT AS A FUNCTION OF TERMINATING RESISTANCE

Variation of the output voltage of the memory core with changes in the terminating resistance was observed in the same 6-wrap Mo-Permalloy core described in the foregoing. The plot in Fig. 16 on semilog scale, beginning with very low voltages for small values of terminating resistances, passes through a fairly straight region between 70 and 700 ohms and then, asymptotically, approaches the open-circuit voltage. The halfway point, or point of maximum power transfer, is reached near 200 ohms. A common way of specifying impedance of a winding on a magnetic core is in the unit "ohms per turn squared." For our 20-turn sensing winding, its apparent 200-ohm internal impedance gives, in these units, 1/2 ohm per turn squared. The quadrature-current pulse was 0.64 ampere maximum, about 0.25 microsecond long at the base.

OUTPUT VOLTAGE AS A FUNCTION OF THE AMOUNT OF MAGNETIC MATERIAL IN THE MEMORY CORE

It soon became apparent that the output of a given magnetic core varies a good deal according to its size or the amount of material used. Moreover, the variations seemed to follow different patterns, depending on whether externally or internally generated β fields were used for the readout.

Externally Generated Quadrature Field

One set of metallic-tape cores was ordered containing 1, 2, 4, 8, 16, 32, and 64

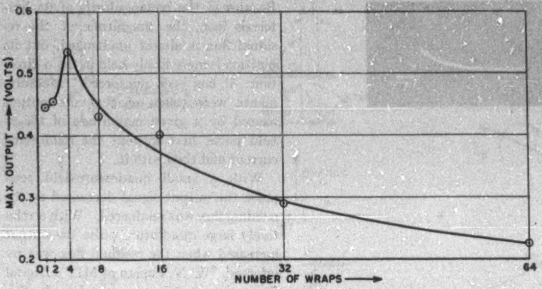


Fig. 17 (above). Output voltage as a function of the number of wraps (external β field)

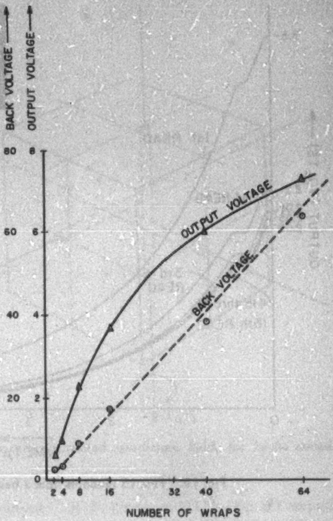


Fig. 18 (right). Output voltage and back voltage as a function of the number of wraps (Internal β field)

wraps respectively. A Ferramic-*D* core with a 10-turn exciting winding was prepared by grinding a slot into it just wide enough to receive the metallic memory cores. The latter were Mo-Permalloy 1/4-mil thick and 1/8-inch wide, with 20-turn sensing windings. Then, with a constant β-field pulse of 0.72 ampere and 0.2 microsecond long, maximum output was recorded for each core. Fig. 17 shows the maximum output for each core in that series plotted against the number of wraps in each. The result seems odd at first; one would expect to find an increased output as the amount of material in the core is increased. But it should be realized that there exists an air gap between the edges of the β-field magnet and the memory core material of at least the thickness of the two flanges of the bobbin on which the metallic tape is wound. The β-field flux is therefore fairly constant no matter how many wraps of material are inserted; however, the flux density will be smaller the more wraps of material are on the memory core. Below four wraps, leakage or saturation (or both) reverses the normal trend.

Internally Generated β Field

The curve of output versus number of wraps of material for the internal β field, the solid line in Fig. 18, starts out quite linearly but soon begins to rise more slowly. Presumably, above a certain large number of wraps no further rise in output would be noticeable as the amount of material is increased.

The back voltage versus number of wraps (broken line) indicates a linear relationship between these two variables, as had been expected. Back voltage in all cases amounted to several times the output voltage.

BACK VOLTAGE AND POWER CONSIDERATION

When considering the nondestructive read system from the point of view of power economy, one is immediately aware of its very low efficiency as a power transformer.

An indication of the peak power involved may be had by multiplying the current peak by the peak of the back voltage. These peaks, from observation, occur simultaneously. Peak power output will then be given by

$$P_{out} = \frac{(E_{out})^2}{R}$$

where E_{out} is the peak voltage appearing across the sensing winding and R is the terminating resistance. Four typical cores are illustrated in Table I; setups *L* and N_{02} used internally generated β fields, setups *C* and M_{04}, externally generated β fields.

The materials used in Table I are:

Setup C (shown in Fig. 9)
Memory core: *MF-1118* size *F-259* ferrite core, with 20-turn sensing winding.
Quadrature-field magnet: *MF-666* size *F-262* with 20-turn exciting winding.

Setup M_{04}
Memory core: 4-wrap 1/4-mil-thick 1/8-inch wide Mo-Permalloy, 20-turn output winding.
Quadrature-field magnet: Ferramic *D* 13/16-inch outside diameter with 10-turn exciting winding.

Setup *L* (shown in Figure 11)
Mo-Permalloy core 1/4-mil thick, 1/8-inch wide, 10-wrap, with 20-turn sensing winding.

Setup N_{02}
Mo-Permalloy core 1/4-mil-thick, 1/8-inch wide, 2-wrap, with 20-turn sensing winding.

Table I. Approximate Efficiency Calculations

	Setup			
	External β Field		Internal β Field	
Peak	C	M_{04}	L	N_{02}
I_{Beta} (ampere)	1.0	0.72	0.72	0.168
V_{Back} (volt)	130	39.0	3.5	0.5
P_{in} (VA)	130	28.1	2.52	0.084
E_{out} (volt)	0.55	0.6	0.6	0.1
R (ohm)	93	93	100	93
P_{out} (watt)	0.0033	0.0039	0.0036	0.00011
Ratio: $\frac{P_{in}}{P_{out}}$	39,000	7,200	700	760

As may be seen, back voltages vary a great deal. In the four cases shown they range from 130 to 1/2 volt.

For applications where outputs of 0.1 volt would be sufficient and where small back voltages are desirable, setup N_{02} might be suitable. For applications where large output voltages are needed, a 120-wrap Deltamax core might be useful. Open-circuit outputs of over 50 volts in a 20-turn sensing winding have been obtained with a fast-rising 3.2-ampere β-field-current pulse. This same setup yielded over 20 volts into a 93-ohm terminator.

TRANSIENT BEHAVIOR

An important question arose early in the investigation: how permanent is this nondestructive read system? It was ob-

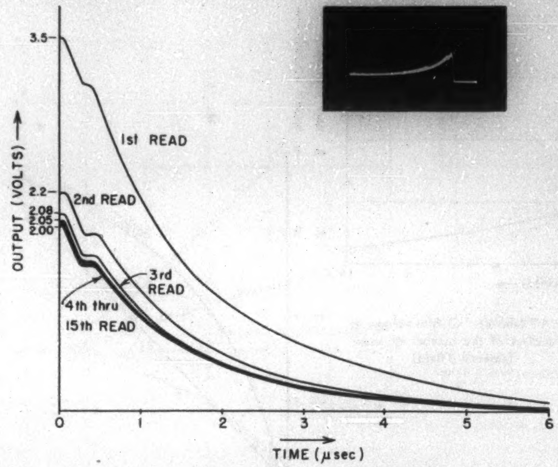


Fig. 19. First 15 readouts from a freshly magnetized core

served that for a quadrature-field pulse of given magnitude and length, the readout signal did not always remain constant. With relatively short pulses of small amplitude, no change could be detected; but when the quadrature-field pulse approached the limit of the nondestructive system, a definite change could be observed in the magnitude of the readout signal. From a maximum output obtained with a freshly magnetized core, the output would shrink, rapidly at first and then more slowly, until it finally settled at some steady value.

Fig. 19 shows the traces, as observed on the oscilloscope, of the first 15 nondestructive readout signals obtained from a freshly magnetized core with internally generated β-field pulses. (The actual photograph is reproduced in the insert; for better visibility the traces were redrawn to a much larger scale.) The pulse was 0.36 ampere high and 6.5 microseconds long, which, for the type of core used, is very near the destructive limit.

The transient behavior appears to be of an exponential nature. The values obtained are given in Table II. The core used was a 1/4-mil-thick 1/8-inch-wide 40-wrap Mo-Permalloy core. Repetition rate was 4 kc. At this relatively slow rate, the steady state was reached after about 2 to 3 seconds; for over an hour thereafter no further changes were observed. Other endurance tests were run up to 1,200,000,000 readouts, and in all cases no further deterioration of the output signal was noted once it had settled down. The number of pulses needed to reach a steady state seems to be inversely related to the magnitude of the quadrature-field pulse. In several cases involving short pulses (0.2 microsecond), steady state is reached after the second reading.

READOUT TIME

As far as was observed, the lower limit of the readout time seems to depend only on the rise time and length of the β-field-current pulses which the available equipment is able to provide. The fastest pulses so far have been of 0.1-microsecond rise time and 0.2-microsecond total length. As illustrated in Figs. 13 and 14, the readout signal occurs almost simultaneously with the quadrature-field pulse. There seems to be every reason to expect, if even steeper and narrower quadrature-current pulses could be provided, that the readout signal would also become steeper and narrower, and the voltage correspondingly higher.

Some Further Experiments

EFFECTS OF REINFORCING RESIDUAL FLUX

An investigation was made to see what happens when direct current is allowed to flow in the memory core's magnetizing winding (α-field winding) while β-field pulses are applied. The effect of the direct current is to increase the α field. Because of the rectangularity of its hysteresis loop, the magnitude of the residual flux is almost unchanged, but its position is more firmly held in the α direction: it has been reinforced. Measurements were taken of the core output caused by a given magnitude of the β-field pulse, first without the reinforcing current and then with it.

With a small quadrature-field read pulse the output signal decreased as the residual flux was reinforced. With a relatively large quadrature pulse the output increased when the residual flux was reinforced. W. N. Papian of MIT's Digital Computer Laboratory explained this phenomenon as follows.

As the small β field causes the residual-flux vector to rotate, the sensing winding sees a certain flux change, $\Delta\phi$ in Fig. 20 (A). When the residual flux is reinforced, the vector is still of the same magnitude but can no longer be rotated as easily as before, Fig. 20(B), resulting in a smaller flux change and hence in a smaller voltage output.

The large β-field pulses, in the absence of direct current in the magnetizing winding, are powerful enough to switch some of the smaller domains into the 90-degree direction, Fig. 20(C), and the residual flux in the α direction reaches a steady state somewhat below its saturation value. The domains still in the α direction are deflected as usual; again, the sensing winding effectively sees a change of flux, $\Delta\phi$. Now, when sufficient direct current is allowed to flow in the magnetizing winding, the small domains are reswitched into the α direction, and thus the total residual flux in that direction is larger than before. If the rotation caused by the β-field pulses is not too much reduced, the sensing winding will see a larger change of flux than before.

SWITCHING TIME OF A MEMORY CORE WHILE β FIELD IS APPLIED

The primary function of the β field is to rotate the remanent flux away from its easy direction of magnetization, that is,

Table II. Successive Output Voltages

Read Pulse		Output (V)
1st		3.5
2nd		2.2
3rd	(estimated 2.08)	2.0–2.1
4th	(estimated 2.05)	2.0–2.1
15th		1.9–2.0
100th		1.6–1.7
10,000th		1.5
15,000,000th		1.5

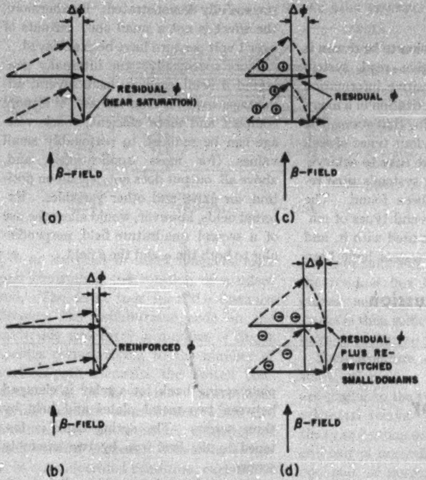


Fig. 20. Effect of reinforced residual flux

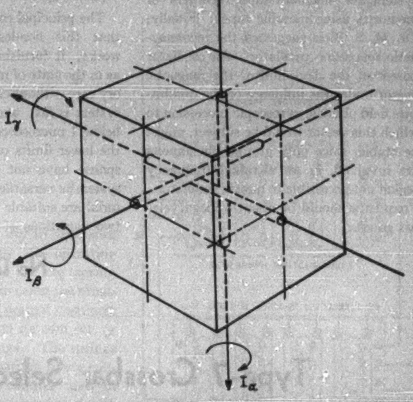


Fig. 21. Internally generated quadrature fields for ferrite material

away from its most stable state. It was thought that the application of a β field while the α direction was being switched, say, from a ZERO to a ONE would facilitate this switching process and thereby reduce the switching time of the core. Indeed, it was found that switching time could thus be reduced by up to 30 per cent for a given number of ampere-turns.

REGENERATIVE EFFECT

During the attempt to find the maximum amplitude and length of the quadrature pulse which various memory cores would tolerate and still yield a nondestructive readout signal, an unexpected phenomenon was observed: under certain circumstances, when information retained by a memory core had been largely destroyed by the application of too powerful a quadrature field, this information could be restored merely by shorting the sensing (output) winding of the memory core during the application of a few quadrature-field pulses. For simplicity's sake, this effect will be referred to as the regenerative effect.

The effect is by no means found in all cases. So far, it has been observed in only four of the metallic cores, out of over 30 which were used during this investigation, and in none of the ferrite cores.

In a typical case, a 40-wrap 1/4-mil-thick 1/8-inch-wide Mo-Permalloy core with a 20-turn unterminated output winding was used. With a 0.5-ampere current pulse of 0.1-microsecond rise time and 0.2-microsecond length, a ONE output of 13.5 volts was obtained. As the current-pulse length was gradually increased, the output fell, rapidly at first and then more slowly, until at 50-microsecond pulse length only 1.4 volts was measured. The pulse length was then gradually reduced back to 0.2 microsecond; the output, however, did not rise above 2 volts, indicating that the information had been largely destroyed, probably because most of the domains had switched into the 90-degree direction of the β field.

The output winding was then short-circuited while the β-field pulses continued, and regeneration took place; i.e., after the short circuit across the output winding had been removed, the output voltage again stood at 13.5 volts, suggesting the remanent flux had been restored to its original magnitude.

When instead of a short circuit a small resistor was placed across the output winding, the gradual regeneration could actually be observed on the oscilloscope. Similar results were obtained when the core was magnetized along minor hysteresis loops: application of quadrature field pulses to the core, while its output winding was short-circuited, brought the remanent flux back to its full saturation value. No satisfactory explanation of this phenomenon has yet been found. Any thoughts the reader might have about it would be appreciated.

INTERNAL β FIELD FOR CERAMIC CORES

A system for generating a β field, and even a γ field, internally, was proposed by R. F. Jenney of MIT's Digital Computer Laboratory. A small cube of ferrite material, Fig. 21, is drilled in the three perpendicular directions in such a manner that the three hollows thus obtained pass near one another without actually intersecting. The cube may then be regarded as a core, with the α-current (I_α) conductor representing the magnetizing winding and the β- and γ-current conductors carrying the quadrature field currents.

Tests with such a drilled-out block of ferrite *MF-1326B* show that nondestructive reading can be obtained by this scheme.

Some Possible Applications

Applications for the nondestructive read system could be found wherever permanent or semipermanent memories are used in computers or similar machines. A magnetic-core binary counter (currently being investigated at the MIT Digital Computer Laboratory) could be equipped with this nondestructive read system, so that the actual count at any one instant could be ascertained without interruption of the process. A permanent register could be built consisting of, say, 16 cores which might be set up by individual toggle switches. If their quadrature fields could be pulsed in series, or possibly if they could all be disturbed by the same quadrature field, all 16 cores could yield their information simultaneously, without the necessity of having to rewrite the information.

Another possibility might be the use of

internally generated β field in a static read system for magnetic-tape recorders or memories using metallic tapes. Finally, Dr. M. S. Blois suggested the ferromagnetic relaxation oscillator. An oscillator based on the deviation of the magnetic vector under the influence of the quadrature field plus the gyroscopic precession to which this vector would be subject, might be stable, since only atomic phenomena are involved, in an ultrahigh-frequency region not accessible to quartz resonators. Frequency should be about 5 megacycles per gauss.

Conclusions

The principal conclusion to be drawn is that this nondestructive read system works; it furnishes positive information as to the state of magnetization of a memory core without changing that state. The system is fast, and readout times of well below 1 microsecond are easy to achieve; the lower limits of the system's time response have not yet been found. The system is versatile: several types of material are suitable to be used with it, and two variations of the system have been successfully demonstrated. Furthermore, the effect is not a small one; outputs of over 1 volt per turn have been achieved.

More specifically: the internally generated β field seems to have several advantages over the external one: it is more compact and more efficient; back voltage can be reduced to reasonably small values (for mass applications); and, above all, output does not depend on position, air gaps, and other variables. External fields, however, would allow the use of a second quadrature field, perpendicular to both the α and the β field.

No Discussion

Type-7 Crossbar Selector

ROBERT W. HUTTON
NONMEMBER AIEE

THE TYPE-7 crossbar selector is a mechanism for selecting one of several trunk groups in a dial telephone switching system. Each of these trunk groups may lead to a subsequent rank of group selectors, to connectors (final selectors), to outgoing trunks, or to one of several special services. Switching in this case is based on a system of decimal numbering; consequently the selector, in response to a single pull or digit of a telephone dial-calling device, provides a choice of one trunk group out of 10. The selector consists of a group of relays, a magnetic impulse counter to register the digit, and certain permanently associated portions of several crossbar switches. Generally, each selector has access to 100 outlets, which are arranged in 10 groups of 10 paths each.

In modern telephone practice, it is widely recognized that to provide trunking flexibility, group selectors are desirable, and often a necessity, even in small initial installations, such as a community dial office. The ease of equipment growth thus gained is self-evident. With the advent of nation-wide toll dialing and customer toll dialing, the area and exchange numbering plans frequently demand the use of so-called 2–5 numbering in the exchange, that is, 2 letters followed by 5 numerals. It is generally more practical and economical to accomplish this by the use of group selectors at the outset, at the same time providing for an orderly growth beyond an initial equipment of, say, 60 or 80 line terminals. The fundamental operation of the type-7 selector and the division of available outlets into groups of 10 are based on well-known practices. However, the use of crossbar switches for establishing the selected path on a direct-access basis involves several details of operation that differ somewhat from those generally used in decimal systems. The following descriptions and illustrations are presented for the engineer and technician who are interested in the circuits, apparatus, and equipment arrangements employed.

Elements of the Selector Equipment

The type-7 selector uses the 4000-type relay, which consists of a spring bank assembly, heel iron, relay coil, and armature assembly as shown in Fig. 1. The spring bank assembly consists of contact springs, insulators, lifting cards (or ladders), and clamping plates. The contact springs have conventional terminals for soldered wire connections. The complete spring bank for a relay is clamped between two metal plates and held by three screws. The spring bank is fastened to the heel iron by two assembly screws.

The winding of the relay coil is made up of layers of enameled magnet wire with a sheet of cellulose acetate between adjacent layers. When the winding operation is completed, the protruding ends of the interleaved cellulose-acetate sheets are coalesced to seal the coil against entry of moisture.

The relay coil and spring bank are assembled to a conventional L-shaped heel iron, to which the armature assembly is also attached.

People in the communications field have long understood the possibility of completing an electric path by connecting two crossbars where they intersect. For example, the peg-type telegraph switchboard, which is shown in Fig. 2, has been in use for more than 50 years. It should be noted that, with a considerable number of lines, it is possible to connect any two of them together by connecting both to a commonly accessible link. For example, as shown in Fig. 2, *L*-3 is connected to *L*-12.

The crossbar switch differs somewhat from other types of telephone switches, in so far as direct electric paths are established through precious-metal contacts with only two rapid relaylike operations. Associated apparatus that assists in making the connection is subsequently released for further calls. The crossbar switch, Fig. 3, is composed of a formed metal frame on which is mounted a multiple bank comprising a number of horizontal and vertical paths, the cross-point selecting and establishing mechanisms, and magnets.

Located at the ends of the switch are

Paper 53-408, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Kansas City, Mo., November 2–6, 1953. Manuscript submitted August 11, 1953; made available for printing September 23, 1953.

ROBERT W. HUTTON is with the Kellogg Switchboard and Supply Company, Chicago, Ill.

The author acknowledges the contributions of other engineers in the development of the type-7 crossbar selector, particularly C. A. Adams, who did much of the basic work, and A. T. Sigo and K. L. Liston, whose consultation and encouragement contributed greatly to the project.

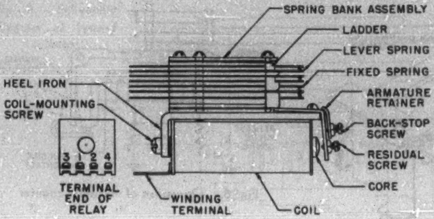


Fig. 1 (left). A 4000-type relay

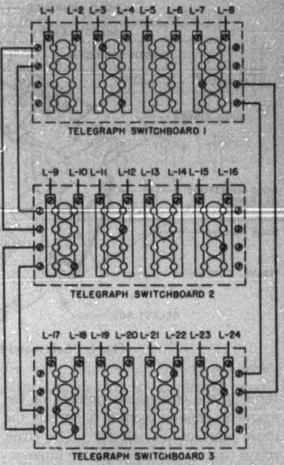

L-1 L-2 L-3 L-4 L-5 L-6 L-7 L-8
TELEGRAPH SWITCHBOARD 1
L-9 L-10 L-11 L-12 L-13 L-14 L-15 L-16
TELEGRAPH SWITCHBOARD 2
L-17 L-18 L-19 L-20 L-21 L-22 L-23 L-24
TELEGRAPH SWITCHBOARD 3

Fig. 2. Peg-type telegraph switchboard

The lines are connected to vertical conducting strips and the circular conducting islands are connected in horizontal rows. Metallic pegs make contact between the cross bars

five or six pairs of select magnets, each pair controlling the rotation of a select rod. The select rods span the distance between the switch-frame ends on the underside and carry a number of finger springs corresponding to the number of vertical-path elements the switch contains. Fig. 4 shows how these finger springs actuate select levers, which are mounted on the armature assembly of each vertical element. When the switch is in an unoperated condition, each select lever rests between two adjacent ladders. When a select magnet is energized, the select rod rotates a few degrees causing the finger springs to move each of the select levers under one of its adjacent ladders. The direction in which the select lever moves depends on which of its two controlling magnets is energized. When the select-rod action has taken place, Fig. 4, the hold magnet of the appropriate vertical path is energized, operating its armature and causing the select lever to lift the ladder, closing the contacts of the selected vertical and horizontal paths. The select magnet is then released, restoring the select rod and all other finger springs to their normal positions. The switch is then available for another selection. The hold-magnet armature remains operated until the subscriber releases the circuit.

The magnetic impulse counter pictured in Fig. 5 is fundamentally a 2-step relay. However, instead of the usual single armature it has 10 armatures. The armatures operate in sequence, each one responding to its respective pulse of current in the driving coil. Once operated, the armatures are held by the residual magnetism in the core. The counter is shown schematically in Fig. 6.

The counter-coil has in addition to its driving or pulsing winding, a second winding that is used to bring about the release of any operated armatures. Release is effected by passing direct current through the release winding in the proper direction to induce into the core a magnetic flux of opposite polarity to that induced by the operating pulses. The release flux is of sufficient intensity to neutralize the residual flux left by the operating pulses, and the tension of the spring stacks is then sufficient to cause the armatures to restore to their normal positions.

Each armature acts on its own set, or stack, of contact springs. The springs are similar to the type used in many conventional relays. The counter used in the type-7 group selector is equipped with one pair of normally closed contacts and one pair of normally open contacts in each of its 10 spring stacks.

Equipment Arrangement

The equipment used in the group-selecting stage of a type-7 office, shown diagrammatically in Fig. 7, consists of the selector relay units, crossbar switches and the common test circuits. Each selector relay unit has a group of relays and a magnetic impulse counter that perform the usual functions of digit registration, holding, etc. Also, each selector relay unit is permanently wired to a horizontal path of one or more crossbar switches, depending on the number of outlets required.

A test relay circuit is provided to serve each 12 selector relay units. It is capable of testing all paths in the selected group of outlets simultaneously, choosing the idle path of the lowest number, and directing the crossbar switch to establish a connection to the chosen outlet. Such operation permits the trunks individually accessible to selectors of a group to be taken into use whenever available, reserving those commonly accessible to a greater number of selector groups for use when the individuals are all engaged.

Operation of the Selector

Operation of the typical type-7 selector circuit is illustrated by Fig. 8. The calling party lifting the telephone handset causes a selector to become associated with his line by way of a line finder. The selector relay unit returns dial tone to the calling party, indicating that it is prepared to receive signals from the telephone dial. The calling party dials the first digit of the number, and the cut-through action of the selector is accomplished in the following manner.

The series relay, which operates on the first pulse, will release after the last pulse of the digit is received, extending ground

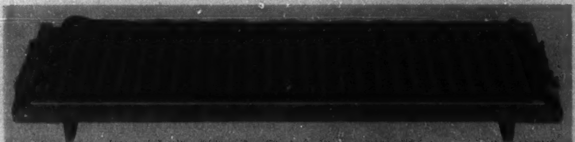


Fig. 3. Crossbar switch

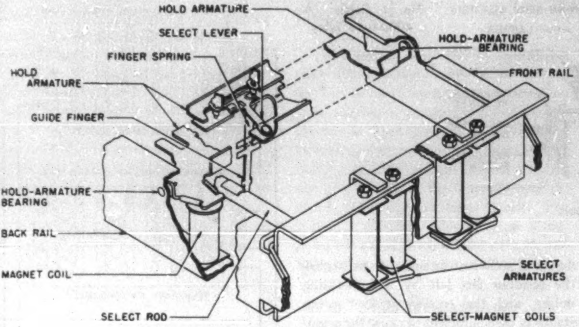

HOLD ARMATURE
SELECT LEVER
FINGER SPRING
HOLD ARMATURE
GUIDE FINGER
HOLD-ARMATURE BEARING
FRONT RAIL
HOLD-ARMATURE BEARING
BACK RAIL
MAGNET COIL
SELECT ROD
SELECT ARMATURES
SELECT-MAGNET COILS

Fig. 4. Select-and-hold magnet mechanism

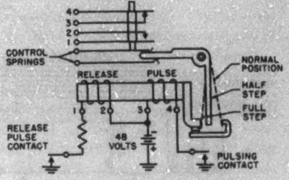

4
3
2
1
CONTROL SPRINGS
RELEASE
PULSE
NORMAL POSITION
HALF STEP
FULL STEP
1
2
3
4
RELEASE PULSE CONTACT
48 VOLTS
-
+
PULSING CONTACT

Fig. 6. Operation of magnetic counter

through the off-normal contacts of the magnetic impulse counter to operate the chain relay of the selector. The chain relays (one in each selector) are interconnected so that only one can be operated at a given time. The operation of the chain relay connects ground through the selected path of the magnetic impulse counter to operate the call-in relays of the trunk group, corresponding to the digit dialed. These relays connect the sleeve leads of the selected trunk group to the common test relays, and prepare the select-magnet and hold-magnet operate paths to the associated crossbar switch. The chain relay also closes ground through contacts of the select magnet operate relay to energize a select magnet of the crossbar switch. Since the selector relay circuit involved in the call being considered is permanently associated with a specific horizontal path in the crossbar switch, it is necessarily associated with a particular select magnet of the switch.

The trunk sleeve conductors, which have been "called in," are connected to the test relay circuit. There usually are 10 test relays, so arranged that only one of them can be fully operated at a time. They are adjusted so that they will operate in series with 250 ohms to negative battery (indicating an idle-path condition); however, two relays testing the same sleeve wire cannot operate because of the current-limiting effect of the 250-ohm resistor. This prevents double selection of trunks on simultaneous testing by two selector groups. Operation of any common test relay chooses the path to the next stage. The select magnet operating closes off normal ground through contacts of the operated test relay to energize the corresponding hold magnet of the crossbar switch.

Operation of the hold magnet of the crossbar switch closes the selected crosspoint, and in turn grounds the sleeve to operate the cut-off relay in the selector relay unit. The cut-off relay releases the chain relay, which in turn releases the common test relay circuit. The hold magnet locks to ground on the sleeve and the loop of the calling line is extended through the operated crosspoints of the crossbar switch to the selected circuit in the next stage. Operation of the selector cut-off relay releases the line and hold relays of the selector, but the next-stage circuit returns ground on the sleeve conductor to maintain the established path. The magnetic impulse counter remains operated until the selector cut-off relay restores on release of the connection.

While at a given instant, only one call in a group of 12 selector circuits may be in the process of actual switching, the operation is so rapid that a single test circuit is well able to perform the tasks for the group. Only after the selecting digit has been recorded on its counter, is the selector permitted to demand the test circuit. The test circuit acts in a fraction of a second and, having completed the switching operation for that call, becomes available for the next demand which may come from any other selector circuit in the same group.

If the common test relay circuit finds that all trunks are busy, the busy relay of the selector relay unit will be operated. Associated common equipment will be released, and busy tone returned to the calling party.

It should be pointed out that the selector just described is the basic one, and that there are variations to meet special requirements. One of these variations is the digit-absorbing selector. The necessity for such a switching device frequently arises because of the technical requirements of nation-wide operator toll dialing, and nation-wide customer toll dialing. Normally, the directory numbers in an exchange include only the digits necessary for the local switching and selecting functions. However, the introduction of long-distance dialing, with the nation-wide numbering plan, necessitates in

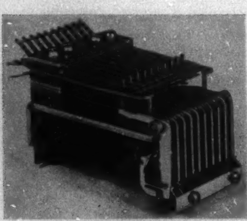


Fig. 5. Magnetic impulse counter

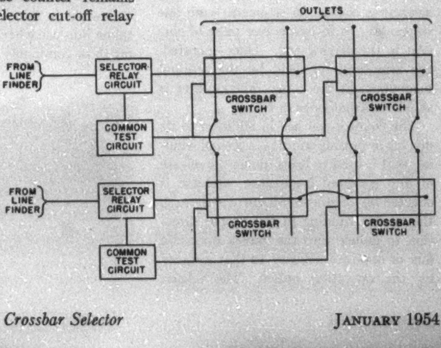

OUTLETS
FROM LINE FINDER
SELECTOR RELAY CIRCUIT
COMMON TEST CIRCUIT
CROSSBAR SWITCH
CROSSBAR SWITCH
FROM LINE FINDER
SELECTOR RELAY CIRCUIT
COMMON TEST CIRCUIT
CROSSBAR SWITCH
CROSSBAR SWITCH

Fig. 7 (right). Block diagram of type-7 selector

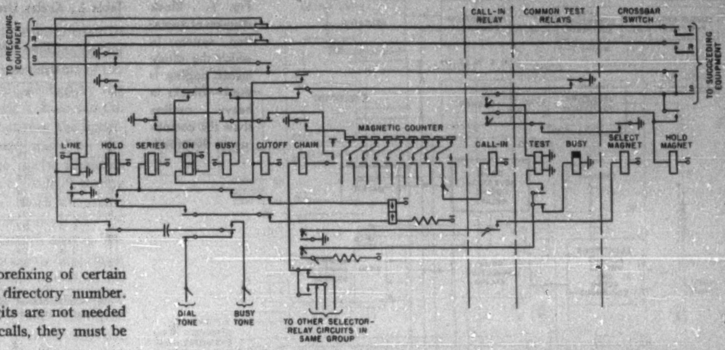


Fig. 8. Schematic diagram of type-7 selector

some exchanges the prefixing of certain digits to the normal directory number. Since these prefix digits are not needed for locally originated calls, they must be absorbed.

The absorption of unneeded prefix digits is accomplished in the type-7 selector by registering them on the magnetic counter in the usual manner and, if the digit is to be absorbed, the counter is released so it will be prepared to receive the next digit. On receipt of a significant digit, the selector action proceeds as in the basic selector already described.

It should be mentioned that the connector (or final selector) of the type-7 crossbar system utilizes the same type of crossbar switches, relays, and magnetic counters used in the group selector. However, the connector normally permits selection of a particular line terminal in a group of 100, rather than choosing an idle path in a selected group of trunks.

Conclusion

While a crossbar switch associated with a selector group may have several of its transmission paths closed at a given time, only one path may be in the process of closing at a given instant. To minimize the time required in testing available paths and in closing the desired crosspoints, the type-7 selector has been arranged to test all available outlets of the selected group simultaneously rather than in sequence. By this method, the testing and connecting time is no greater for the last-choice path than for the first-choice path. These factors made possible the design of a direct-access decimal-type crossbar group selector that readily cooperates with decimal systems of other types.

No Discussion

The Diode Matrix as a Component in Relay Switching Circuits

G. L. BUSH
MEMBER AIEE

THE diode matrix is usually thought of as a component in electronic circuits using diodes of small physical size and having current-carrying capacity in microamperes. This paper describes an application of the diode matrix in a relay switching circuit, where the diodes are small selenium rectifiers with current-carrying capacity sufficient to operate relays. In this application an effort was made to use commercially available selenium-rectifier units for economy in purchasing the small quantity involved.

Statement of the Problem

The problem leading to this development was the introduction of magnetic-drum storage into a stock-exchange quotation system, to distribute bid-ask quotations to brokers in the vicinity of the exchange. It is an automatic system through which the broker receives the bid-ask prices on a stock ticker by dialing a 3-digit code number assigned to the stock. The major components of the system are shown in Fig. 1.

The operation of the system can best be described by following the progress of a call through the various blocks of equipment shown in the figure. The broker desiring the current price looks up the three digit code number for the stock in a code-assignment list. When he is ready to dial, he presses a request button on his dial set. The operation of this button connects his dialing circuit and ticker line through the line connecting equipment to one of the 24 transmitters which may be idle at the time. When the connection to the transmitter is complete, a ready lamp lights on the brokers' dial set, telling him that the equipment is ready to receive his dialing. The operation thus far is similar to the operation of a telephone system when the subscriber picks up the hand set and receives a dial tone. The dialed code-number digits are stored in the transmitter while waiting access to the magnetic drum storage unit where the desired price information is stored. The transmitter also contains the ticker transmitting equipment and the dialed digits stored in the transmitter are sent back to the

---

Paper 53-403, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Kansas City, Mo., November 2-6, 1953. Manuscript submitted March 23, 1953; made available for printing August 17, 1953.

G. L. BUSH is with the Teleregister Laboratories, New York, N. Y.

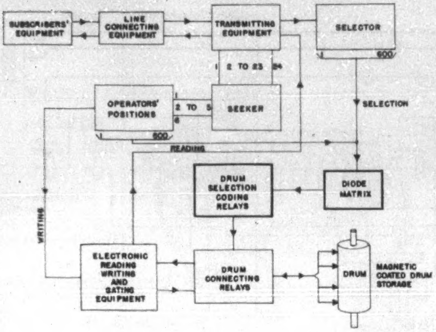

SUBSCRIBERS' EQUIPMENT
LINE CONNECTING EQUIPMENT
TRANSMITTING EQUIPMENT
SELECTOR
1
600
1 2 TO 23 24
OPERATORS' POSITIONS
2 TO 5
6
SEEKER
SELECTION
1
600
READING
DRUM SELECTION CODING RELAYS
DIODE MATRIX
WRITING
ELECTRONIC READING WRITING AND GATING EQUIPMENT
DRUM CONNECTING RELAYS
DRUM
MAGNETIC COATED DRUM STORAGE

Fig. 1. Block diagram of quotation system in which the diode matrix coding is used. Blocks in heavy outline show the components described

Table I. Codes Used in the Coding Circuit

Drum Position Selection					
5 Unit Code				4 Unit Code	
No.	Code	No.	Code	No.	Code
1	1	14	3-5	1	1
2	2	15	4-5	2	2
3	3	16	1-2-3	3	3
4	4	17	1-2-4	4	4
5	5	18	1-2-5	5	1-2
6	1-2	19	1-3-4	6	1-3
7	1-3	20	1-3-5	7	1-4
8	1-4	21	1-4-5	8	2-3
9	1-5	22	2-3-4	9	2-4
10	2-3	23	2-3-5	10	3-4
11	2-4	24	2-4-5		
12	2-5	25	3-4-5		
13	3-4				

Note: Combination codes used in the 1 out of 25 and 1 out of 10 selections.

brokers' ticker during this waiting period as a check of his dialing.

The magnetic-drum storage equipment is time shared between the 24 transmitters and the 6 operators' positions through the seeker. The purpose of the operators' positions is to keep the prices stored on the magnetic drum up to date with the trading. The seeker is a relay-switching device which connects the next transmitter or operators' position awaiting access to the magnetic-drum storage equipment at the end of the operation of fulfilling the previous request. When a transmitter gets access to the magnetic-drum storage, the 3-digit code number stored in the transmitter is translated by the selector into the energization of one of 600 single-wire selection leads previously used to connect the transmitter to a section of the display-board storage, where the "bid" and "ask" prices of the stocks are stored. In the new system, these leads are coded by use of the diode matrix with the position code of the same information on the magnetic-drum storage. When a transmitter has access to the magnetic-drum storage, the electronic equipment is caused to read the price information stored in the section of the drum selected and store it in the transmitter for retransmission over the ticker line to the brokers' ticker. If an operators' position has been given access to the magnetic-drum storage equipment, it causes the electronic equipment to write the price information stored on the operators' keyset in the section of the magnetic drum selected by one of 100 keys on the operators' keyset.

The components described in this paper are shown in heavy outline on the block diagram. As shown in the diagram, the problem is to provide a magnetic-drum position-selection code for each of the 600 leads used to select an indicator storage

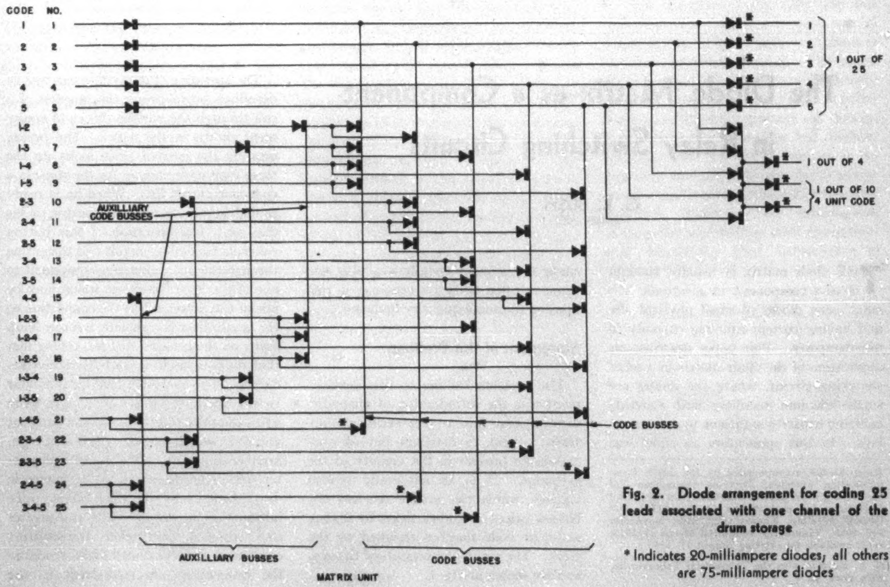

CODE NO.
1 1
2 2
3 3
4 4
5 5
1-2 6
1-3 7
1-4 8
1-5 9
2-3 10
2-4 11
2-5 12
3-4 13
3-5 14
4-5 15
1-2-3 16
1-2-4 17
1-2-5 18
1-3-4 19
1-3-5 20
1-4-5 21
2-3-4 22
2-3-5 23
2-4-5 24
3-4-5 25
AUXILLIARY CODE BUSSES
1
2
3
4
5
1 OUT OF 25
1 OUT OF 4
1 OUT OF 10 4 UNIT CODE
CODE BUSSES
AUXILLIARY BUSSES
CODE BUSSES
MATRIX UNIT

Fig. 2. Diode arrangement for coding 25 leads associated with one channel of the drum storage

* Indicates 20-milliampere diodes; all others are 75-milliampere diodes

in the old system, and now used by both poerators' positions and transmitters to locate the same stored information on the magnetic drum. The requirements of the problem are more closely defined by the character of the magnetic storage and the distribution of stocks between the operators' positions. Storage space on the drum is divided into forty channels around the drum, each capable of storing the prices for 25 stocks. Each operator is equipped to record the prices for 100 stocks on the display boards and on the drum storage. The storage facilities must be made available to the operators for changing the price, and to the brokers for reading out the price. Since the drum storage is arranged to process only one call at a time, the storage facilities must be time shared between the calling transmitters and the operators. This is accomplished through the seeker, which permits only one call to enter the storage at a time. This means that only one of the 600 leads will be energized at a time.

Selection of Codes

For coding purposes, the selection is broken down into three groups, based on the characteristics of the rotary drum storage and the operators' positions. The first group is a selection of 1 out of 25 stocks stored in any one of the 40 channels on the drum storage. The smallest number of code elements to be used, is determined by the value of n in the quantity 2^n, which is larger than 25.

$2^4=16$

$2^5=32$

The 5-element code is therefore the smallest that will have combinations enough to code 25 things. The coding assigned uses all the combinations of the 5 elements of the code taken one, two, and three at a time, as shown in Table I.

Since each operator is equipped to post 100 stocks, or 4 groups of 25, the second group is a 1-out-of-4 selection for the group of 25 in each 100. In this case, so little is gained by use of a permutation code that four separate leads are used in the selection circuit. The third group is a selection of 1-out-of-10 operators positions, provision being made to increase the capacity of the system to 1,000 stocks and 10 operators' positions. The smallest number of elements for coding ten things is four, since $2^3=8$ and $2^4=16$. The codes assigned use all the combinations of the four elements of the code, taken one and two at a time, as shown in Table I, last column.

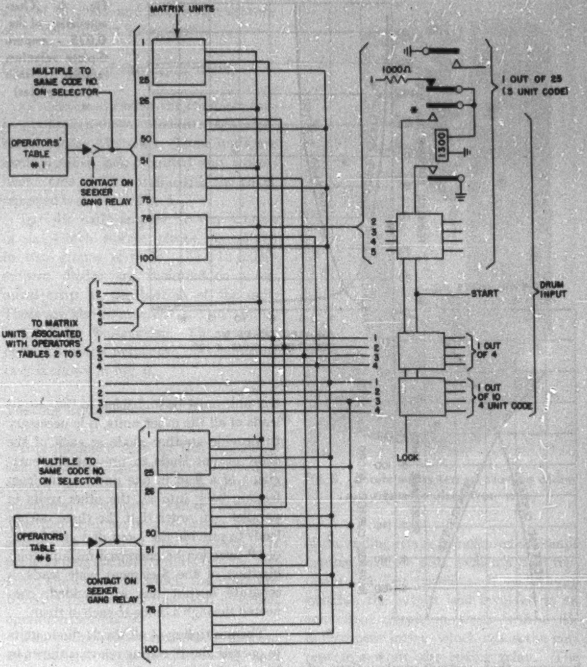

MATRIX UNITS
MULTIPLE TO SAME CODE NO. ON SELECTOR
OPERATORS' TABLE #1
CONTACT ON SEEKER GANG RELAY
1
25
26
50
51
75
76
100
1000Ω
1300
*
1 OUT OF 25 (5 UNIT CODE)
2
3
4
5
START
DRUM INPUT
TO MATRIX UNITS ASSOCIATED WITH OPERATORS' TABLES 2 TO 5
1 OUT OF 4
1 OUT OF 10 4 UNIT CODE
LOCK
MULTIPLE TO SAME CODE NO. ON SELECTOR
OPERATORS' TABLE #6
CONTACT ON SEEKER GANG RELAY

Fig. 3. Multipling of matrix units to form the complete coding of 600 leads

* Indicates preliminary make

The Equivalent Relay Circuit

An effort was made to design a relay circuit to perform the selection operation, using double-winding relays to set up the 5-unit codes for the selection of 1 out of 25. In this circuit 60 single-make relays, 168 double-make relays, and 134 triple-make relays were required. If it were practical to provide relays with three or more windings each sufficient to operate the contact pileup with a reasonable current, the number of relays could be greatly reduced. In the circuit proposed, a winding on each of the 1-out-of-25 selection relays operates in series with a common relay with three make contacts, to mark the 1-out-of-4 and 1-out-of-10 selections.

Building Up the Matrix

As a more economical alternative, it was decided to use a diode-matrix translator. In the buildup of a matrix to perform this coding function, it was found desirable to unitize the selection equipment by providing a separate matrix unit for each group of 25 selection leads. This coding unit has 25 inputs and 8 outputs. The eight outputs provide the five code leads for the 1-out-of-25 selection, one lead for the 1-out-of-4 selection, and two leads for the 1-out-of-10 selection. One of the two leads for the 1-out-of-10 selection is not required for the first four operators' positions, but it is provided in every case to make the units interchangeable.

The design of the matrix itself is based on the quality of a selenium diode to be a reasonably good conductor in one direction and a practical insulator in the other. It is evident that only 1 of the 25 input leads can be energized at a time. In order that the energizating of one input lead will not feed another, every connection from an input to the coding busses must be through a diode. If the matrix is built up on this basis, the number of diodes required will be equal to the sum of the number of times each code element appears in the 25 codes, or 55, see Table I. An examination of the codes shows

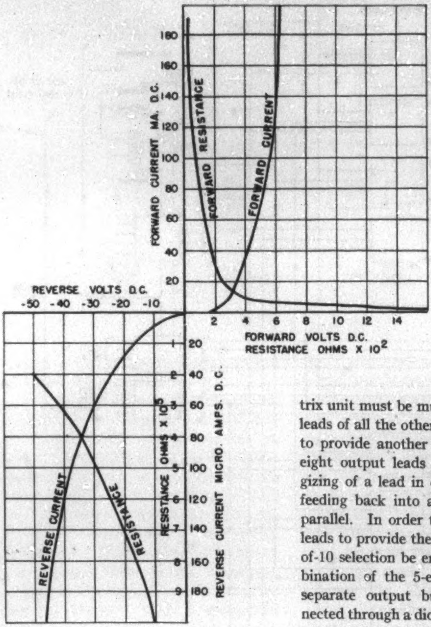


Fig. 4. Characteristics of the 0.075 - ampere 6-plate selenium rectifier (1-inch square plates)

however, that certain combinations of code elements are repeated several times. If an auxiliary bus is provided in the coding circuit for each group of combinations used more than twice, a saving of one diode for each combination more than two can be realized. In the matrix used as shown in Fig. 2, a saving of six diodes was realized by providing busses for the combination groups 1–2, 1–3, 2–3, and 4–5, reducing the total number of diodes to 49.

Since the eight output leads of the matrix unit must be multipled to the output leads of all the other units, it is necessary to provide another diode in each of the eight output leads to prevent the energizing of a lead in one unit matrix from feeding back into all the other units in parallel. In order that the three output leads to provide the 1-out-of-4 and 1-out-of-10 selection be energized for any combination of the 5-element code leads, a separate output bus is provided, connected through a diode to each of them.

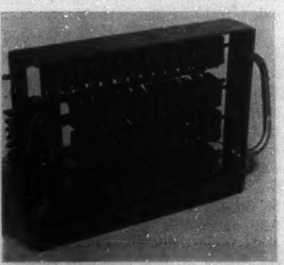


Fig. 5. Diode matrix unit assembly with cover removed

The multipling of all the 24 diode units to operate the 13 coding relays is shown in Fig. 3. In this circuit it will be noticed that the five code leads from each of the 24 matrix units are multipled and connected to the five coding relays for the 1-out-of-25 selection. One output lead of each matrix unit, representing the first 25 stocks of the 100 at each operator's position, is multipled to similar leads from the first matrix unit at the other operators' positions and connected to the no. 1 relay in the 1-out-of-4 selection. Matrix units representing the other groups of 25 are multipled in a similar manner and connected to relays 2, 3, and 4 of the 1-out-of-4 selection group. The remaining two output leads are each multipled for each operator's position, and connected to the four coding relays of the 1-out-of-10 selection, according to the code shown in Table I.

Selection of Diodes

The choice of diodes is based on the relay currents involved, the operating voltage, and the ratio of forward to backward resistance of the diodes to give safe operation. The relay circuit is designed with a preliminary-make locking contact on each relay, so that current through the matrix need be only enough to operate the one spring. The current is further limited in time of application by being applied to the relay through a break contact, which limits the time of application of the current to the operate time of the relay, which is about 0.012 second, see Fig. 3. The operate-readjust current for the preliminary-make spring with the 1,300-ohm coil used is 0.010 ampere. The resistance of the operate circuit is adjusted to give an operate current of about twice this value, for safe operation on 50 volts. This made the total resistance of the operate circuit

$$R=\frac{E}{I}=\frac{50}{0.020}=2{,}500 \text{ ohms}$$

of which 2,300 ohms is in the relay coil and a resistance in series with the operate lead, the remainder being in the matrix. The maximum load carried by any diode in the matrix occurs when a lead with the code combinations 1–2, 1–3, 2–3, or 4–5 is energized. For these combinations, current through the input diode must be sufficient to operate 5 code relays, or $5\times 0.020=0.10$ ampere. The next largest current through a single input diode occurs when the leads with codes 1, 2, 3, 4, or 5 are energized. In this case, four code relays must operate, with a resulting current of $4\times 0.02=0.080$ ampere. In all other cases, the relay load is divided between two or more input diodes, and the current for each will be considerably less.

Since the selenium rectifier will carry considerably more than its average rated current for short periods of time, a smaller diode may be used than would be indicated by the maximum current values

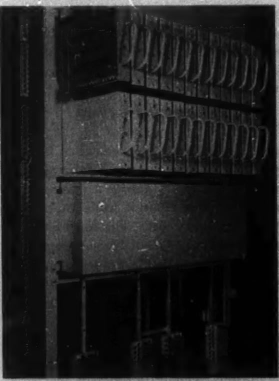


Fig. 6. Mounting of 24 matrix units on the equipment rack

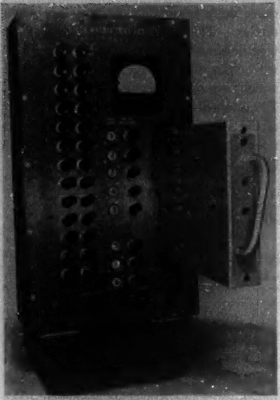


Fig. 7. Diode matrix test set with matrix unit plugged in (front view)

shown above. For this application, commercial rectifiers rated at 0.075 ampere were used, when required to operate two or more code relays. Diodes which carry the current of only one code relay, such as those in the output circuit, are commercial selenium rectifiers, rated at 0.020 ampere. The locations of the 0.02-ampere diodes in the circuit are shown by an asterisk in Fig. 2.

For the operating voltage used (50), rectifiers with only two or three plates would be required. Rectifiers with three plates were tried in this application, but it was found that they could not be made in production quantities with sufficiently high back resistance for safe operation. The rectifiers used were commercial 6-plate units, designed for use on 120-volt a-c circuits. The average characteristics of these units is shown in Fig. 4.

As shown in this figure, the forward resistance for the current used is from 50 ohms to 170 ohms, and the reverse resistance at the 45 volts available across the feedback circuits is about 260,000 ohms. The limiting feedback circuit is that which is trying to operate the fourth and fifth code relays when there are three code relays operated. The feedback path to operate these relays falsely consists of 7 of these reverse circuits in parallel, or 260,000/7 = 37,143 ohms. This gives a current of 45/37,143 = 0.0012 ampere. The current required to just operate the relay is 0.010 ampere. This gives a safety factor against false operation of 0.010/0.0012 = 8.3. The 0.02-ampere diodes are operated at near their rated current, and have an average forward resistance of 70 ohms, and a reverse resistance on 45 volts of over 5 megohms.

Construction of the Equipment

As mentioned earlier the complete matrix is broken down into units each capable of coding 25 stocks. These units are made up into jack-mounted unit assemblies. One of these units with the cover removed is shown in Fig. 5.

The 49 0.075-ampere diodes consist of six 1-inch square plates, assembled in two groups of three. The 13 0.020-ampere diodes are mounted on a terminal strip at the bottom of the unit. These are also commercial 120-volt units, in a tubular construction. The assembly of the 24 matrix units onto the equipment rack is shown in Fig. 6.

Testing Equipment

In equipment such as this, where so many interlocking circuits are involved, it seemed advisable to provide with the equipment adequate testing facilities. This test equipment is shown in Figs. 7 and 8. As shown in Fig. 7, provision is made to plug the matrix units into the test set. The circuit in which the units are tested is the same as that used in actual operation, see Fig. 9.

The coding-relay base, shown plugged into the back of the test set in Fig. 9, is the spare coding-relay base for use in the service equipment. Provision is made on the test panel to energize each of the 25 input leads individually. The operation of the coding relays is indicated on lamps marked with the code elements they represent. A key beside each code lamp switches the output lead involved to an equivalent circuit in series with a 25-milliampere meter, which reads the current to operate the coding relay. The equivalent circuit was necessary because the coding relay opens its own operate circuit before a meter reading can be obtained. The coding relays which operate

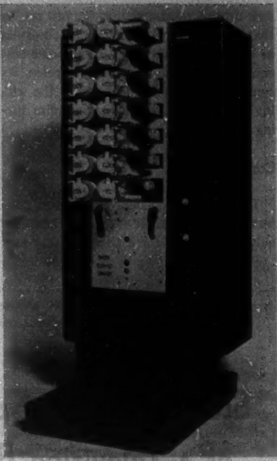


Fig. 8. Diode matrix test set showing coding relay base in place (rear view)

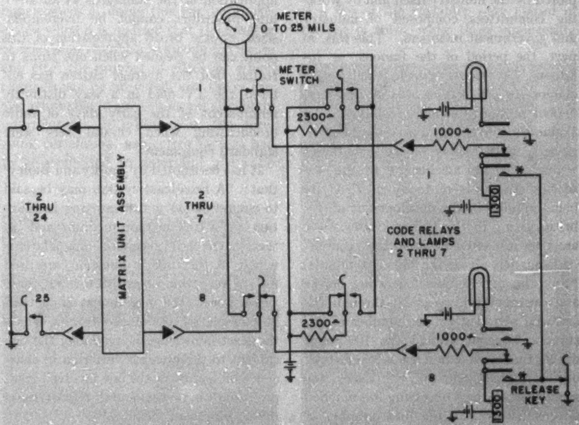


Fig. 9. Schematic circuit of the diode matrix testing equipment

* Indicates preliminary make

lock-up to a release key on the test set, so that the code registered may be observed as long as desired. Even though the test equipment is seldom used in service, its construction was justified to test the wiring and operation of the matrix units during manufacture.

Conclusion

This selection system is designed conservatively, in that it overloads the diode only 25 per cent for a short pulse: only long enough to operate a relay probably not oftener than once every 30 seconds. It is believed that similar circuits can be designed which utilize the instantaneous overload characteristics of the selenium diode to a much greater degree, with satisfactory operation. By permitting an overload of 500 per cent, the small 20-milliampere diodes could be used throughout, with considerable saving in space. By introducing a tube at the input to each relay circuit, it may be possible to use still smaller diodes, but the tubes may require more maintenance than the larger diodes. There is some question as to whether the characteristics of the diode will change with age, especially those which are only occasionally energized. The operation of this system will be a test of the selenium-diode matrix as a tool in relay switching circuits; how good a tool it is only time will tell.

◆

No Discussion

Development of Television Service Standards and Application to Design of a Television Broadcast Network

O. W. B. REED, JR.
ASSOCIATE MEMBER AIEE

IN APRIL 1952, the Federal Communications Commission released its final report and order in television (TV) allocation proceedings which had been instituted in May 1948. In the intervening 4-year period, almost every shred of engineering information and evidence relative to the TV broadcasting service problem was examined and further explored by the industry itself and by working committees composed of industry and government members. This was, in part, the period of the famous, or infamous, TV "freeze" during which public clamor for TV service was at times reflected in Congressional rumblings. An examination of the data used by these engineers, and a study of the hypotheses wrought, yields an insight to the TV service standards of today as they are being applied in the development of TV broadcasting in the United States. We are now just entering upon the proof-of-performance phase for these standards. From the five stations operating before and during World War II, through the artificial period when the number of TV stations hovered around the 100 mark during the freeze period, we have emerged during the past year to a point when the number of stations operating has reached approximately 200, the total number of stations authorized has reached approximately 500, and the present channel assignment table of the Federal Communications Commission provides for 2,000 stations in 1,285 cities and towns.

It will be amply proved before many illustrations are set forth in this paper that the standards as developed should be regarded as guideposts, and carefully qualified. Basic qualifications will be evident in the course of background development and the limitations to the application of the standards as an absolute procedure cannot be overemphasized. Some broad appreciation of this point can be gleaned when one stops to realize that the average citizen has entered the TV field in a way distinctly reminiscent of the early days of radio broadcasting when crystal sets were standard equipment

It has been noted by Jansky and Bailey[1] that: "A broadcast system may be said to consist of (1) a transmitting installation, (2) a transmission medium, and (3) the particular receiving installations which, at the time in question, are used for the reception of programs. Of these three parts, the one least subject to human control, the one concerning which the least knowledge is available, the one subject to the greatest variation in characteristics, and yet the one involved in all problems of governmental regulation is the transmission medium."

The transmission medium in the period 1948–1952 remained as the fountainhead of uncertainties with respect to propagation of signals, which lead various commissioners of the Federal Communications Commission to complain from the bench on many occasions concerning the necessity of allocating frequencies on the basis of inadequate knowledge. One of the basic reasons for this was that frequencies suitable for TV use had but lately been occupied and only a meager amount of measured data was available.

The complexities of this situation were by no means eased by the manner in which broadcasting matured. From even a cursory study of the record, haphazard growth is indicated. The space available for TV today in the region below 300 megacycles (mc) is less than that which was available before the war. Tremendous demands of other services, such as the safety and mobile services, have made it necessary to decrease the number of TV channels in this portion of the spectrum. In 1936, frequency requirements for TV were explored, and in the following year 19 channels were allocated to TV. In 1941, after a lengthy period of experimentation and development culminating in joint consideration by the first National Television Systems Committee, standards were adopted. Provision at this time was made for 18 commercial channels. As a result of further allocation, proceedings commenced in 1944 and, extending over a period of several years, the number of TV channels was reduced to thirteen, twelve of which were to be shared with fixed and mobile services. The TV rules adopted in 1945 contained a table showing the allocation of TV channels to 140 metropolitan districts in the

---

Paper 53-414, recommended by the AIEE Television and Aural Broadcasting Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Kansas City, Mo., November 2–6, 1953. Manuscript submitted July 9, 1953; made available for printing September 1, 1953.

O. W. B. REED, JR., is with Jansky and Bailey, Inc., Washington, D. C.

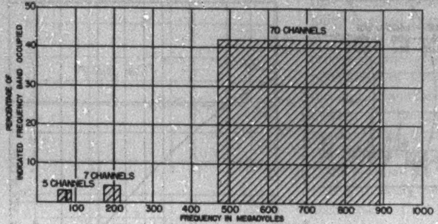


Fig. 1. Distribution of VHF and UHF television channels in the radio frequency spectrum

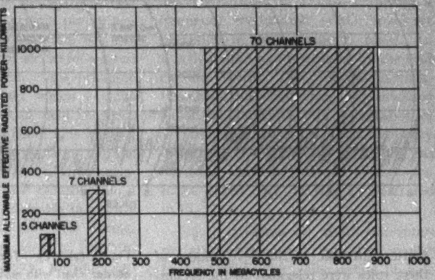


Fig. 2. Maximum allowable effective radiated power in the horizontal plane on VHF and UHF channels

United States. Subsequent to the adoption of these rules, it became apparent that it would not be satisfactory to have frequency sharing between TV and the fixed and mobile services. Therefore, after further hearing, TV channel no. 1 was deleted and reallocated for fixed and mobile use.

In the period of 1948–1952, the necessity was recognized of providing TV service in many cities and towns not covered by the original table. Also, the need was seen for taking into account the deletion of channel 1. These contradictory aims were later to be reflected in the ultimate recognition of the need for more TV channels and additional space in the radio frequency spectrum.

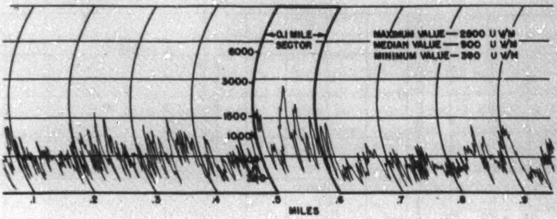
Fig. 1 illustrates the status of TV channel allocations which prevails today as a result of the Commission decision of April, 1952. There are 12 very high frequency (VHF) channels lying between 30 and 300 mc in the spectrum. Five of the channels are between 54 and 88 mc, and seven of them lie in a continuous series between 174 and 216 mc. The ultrahigh frequency (UHF) channels lying between 300 and 3,000 mc in the spectrum number 70. These lie in a continuous band of frequencies between 470 and 890 mc. The TV channels are each 6 mc wide. The carrier frequency for the visual transmitter is located 1.25 mc from the lower frequency edge of the channel, and the carrier frequency for the aural transmitter is located 0.25 mc below the higher frequency edge of the channel. It will be noted that collectively the space in the spectrum allocated for TV use closely approximates one-half the frequency band between 1 and 1,000 mc.

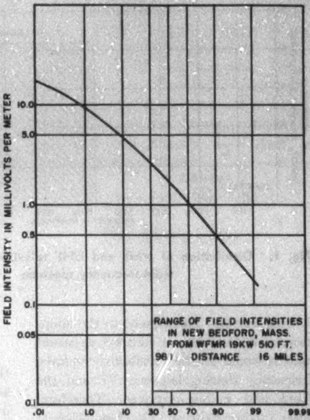
The groundwork for development of an allocation table of assignments to various cities lies in engineering considerations involving numerous assumptions, some of which, as we have noted at the transmitting point and at the receiving point are, within limits, under the control of the developer. At the transmitter, assumptions can be imposed to standardize, over some variable range, effective radiated power and antenna height. At the point of reception, the receiver, the type of antenna to be employed, its average height above the surrounding terrain, and some knowledge of undesired noise levels present can be framed within certain points of reference. Certain assumptions concerning receiver selectivity characteristics which affect adjacent channel responses are of considerable importance. In between these two control points, if we may call them that for a moment, lies as we have stressed, the transmission medium with all its propagation vagaries to be defined over a tremendous frequency range. While it would be tempting to treat each of these three factors as an isolated case, the hopelessness of doing this can be appreciated when it is realized that the number of television channels available for use in any city, the number of cities for which channels can be provided, and the coverage attainable in such cases are all closely interrelated. In addition to the foregoing correlation, subjective factors had to be evaluated to determine the ratio of desired to undesired signals necessary to avoid interference between stations operating on the same and adjacent channels. Studies of this latter subject several years ago resulted in a first introduction of a number of citizens to picture quality factors in TV.

Fig. 3. Field car showing receiving antenna in recording position

Some of the variables at the transmitting and receiving terminals were fixed by the Commission, perhaps arbitrarily, in establishing the present allocation plan. However, with one eye on the present state of the art as regards ability to generate power at the frequencies concerned and the nature of receiving equipment, and the other eye on the transmission medium, the present rules and standards were evolved. At this point we can oversimplify somewhat and say that those matters which were subject to policy decisions resulting in a "fixing" of a few of the variables were placed in the list of rules, while those engineering

Fig. 4. Interior of field car showing field intensity recording and monitoring equipment





matters intrinsic to the development of these rules were relegated to the technical standards category. As an administrative convenience perhaps, the Commission has refused to dignify the Siamese relationship and blankets both these categories into "Rules Governing Television Stations."

Fig. 2, which is similar to the first figure in the abscissa presentation, shows the determination made by the Commission with respect to the maximum allowable effective radiated power to be employed in the three sections of the assigned spectrum. Basic to these determinations were factors affecting reception conditions in the several frequency bands.

A broadcast station produces, at any particular point of reference, an electrical field intensity or signal strength which may be measured by appropriately designed and calibrated apparatus. In broadcasting, it is important to emphasize that it is not the absolute value of field intensity or signal strength produced at a point of reception which determines whether or not the signal is satisfactory, but the ratio of the received intensity to the intensity of other electrical disturbances such as static or automotive ignition and interference from other stations on the same and adjacent channels. Measurements of broadcast signals in the higher frequency bands are most conveniently made by means of mobile field intensity recording arrangements typical of those illustrated in Figs. 3 and 4.

Fig. 5 (above). Typical section of recorder chart with mobile field intensity recorded data

Fig. 6 (right). Probability analysis of mobile field intensity recorded data. Per cent of distance field intensity is greater than indicated level

Fig. 5 contains a section of a recorder chart showing the analysis of one sector of a mobile field intensity recording obtained by use of the equipment illustrated. Within the space of approximately 500 feet it will be noted that the signal varies over a range of approximately seven to one. Several years' experience with measured data of this nature leads one to begin to abandon the use of contours of equal field intensity in terms of which broadcast signals were formerly defined. Summarization of data of the type collected in Fig. 5 by analysis of the sort indicated in Fig. 6 leads to thinking in terms of iso-service contours,

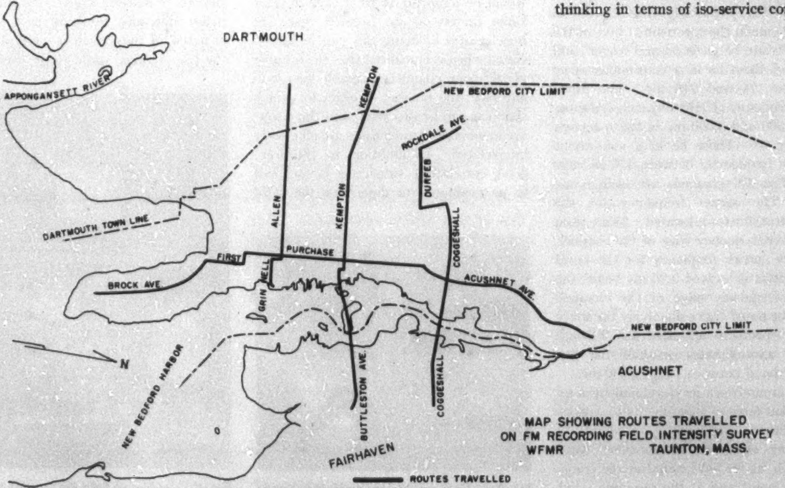

DARTMOUTH
APPONGANSETT RIVER
KEMPTON
NEW BEDFORD CITY LIMIT
ROCKDALE AVE.
DURFEE
ALLEN
KEMPTON
DARTMOUTH TOWN LINE
COGGESHALL
FIRST
PURCHASE
BROCK AVE.
GRINNELL
ACUSHNET AVE.
NEW BEDFORD CITY LIMIT
N
NEW BEDFORD HARBOR
BUTTLESTON AVE.
COGGESHALL
ACUSHNET
MAP SHOWING ROUTES TRAVELLED
ON FM RECORDING FIELD INTENSITY SURVEY
WFMR TAUNTON, MASS.
FAIRHAVEN
ROUTES TRAVELLED

Fig. 7. Map showing routes traveled on recording field intensity city survey

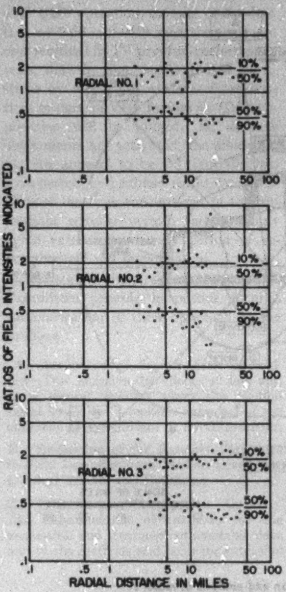


Fig. 8. Ratios and field intensities exceeded for 10 per cent and 90 per cent of the distance to the median value along radials from a transmitter

that is, contours along which the probability is equal that a certain percentage of possible receiving locations will receive a specified grade of service for a given percentage of the time. This principle recognizes another fact which became evident in the course of field intensity studies. Propagation conditions in both the VHF portions of the spectrum and the UHF portions are subject to wide variation with both time and location, and a statistical method is advisable in a study of the service problem. The former practice of never making a standard broadcast field intensity measurement without recording the time of day and date to reconcile ionospheric variations, was carried over into high-frequency work, where the dielectric constant of the air itself became of importance in the transmission medium.

While automatic recordings of the type shown in Fig. 5 are being taken of the field intensity distribution over a city, another receiving set connected to a second receiving antenna may be kept in operation. By this means, it is possible to observe receiving conditions in the car during the progress of each run. Such tests supplemented by several observations at fixed locations provide a base upon which to evaluate the media field intensity required to overcome local noise and interference.

Fig. 7 illustrates the application of the method outlined in a case involving a frequency modulation (FM) broadcast operation in New Bedford, Mass. In this figure the routes traveled by the field car are shown. The propagation vagaries of the radio frequency fields involved would be applicable to interpretation of TV propagation on the lower frequency TV channels. The data in Fig. 6 was compiled from an analysis of measurements taken over the city as illustrated in Fig. 7.

Fig. 8 presents an additional interpretation of recorder chart data similar to that of Fig. 5. The ratios of field intensities exceeded for 10 per cent and 90 per cent of the distance to the median, or 50-per-cent value, give a measure of the variability of the received signal. For this particular set of results it will be noted that the range of variability is independent of the distance from the transmitting station. It will be noted that, while we may be pleased to think in the simplified terms of the median value, it should none the less be kept in mind that for this particular set of circumstances, including frequency, type of terrain, and city characteristics, signal peaks of the order of twice the median value will frequently be experienced and, what is more important, signal minimums of the order of one-half the median signal will occur over any reasonable sector length.

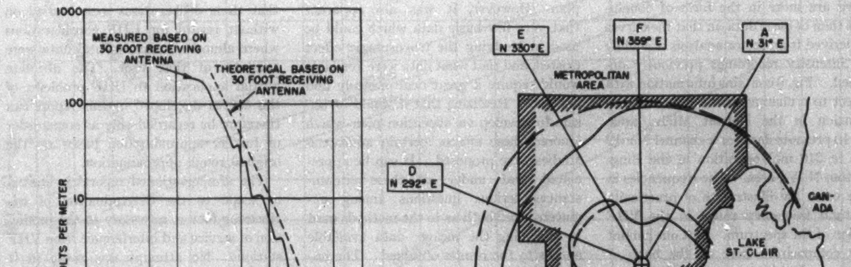

1000
MEASURED BASED ON 30 FOOT RECEIVING ANTENNA
THEORETICAL BASED ON 30 FOOT RECEIVING ANTENNA
100
10
1
.1
0.1
MILLIVOLTS PER METER
FIELD INTENSITY RADIAL
NO. RADIAL A DIRECTION N 31° E
STATION WJBK-TV FREQUENCY 55.25 MC
POWER 16.5 KW ERP DATE 4-27-49
ANT. HT. 485' CHANNEL 2
.1
.5
1
5
10
50
100
DISTANCE IN MILES
E N 330° E
F N 359° E
A N 31° E
METROPOLITAN AREA
D N 292° E
C N 248° E
B N 202° E
CAN-ADA
LAKE ST. CLAIR
5.0 MV/M
CANADA
0.5 MV/M
0.5 MV/M
KEY
THEORETICAL
MEASURED
SCALE OF MILES
10
0
10
20
30

Fig. 9 Theoretical versus measured contours for a TV broadcast station and analysis of radial A

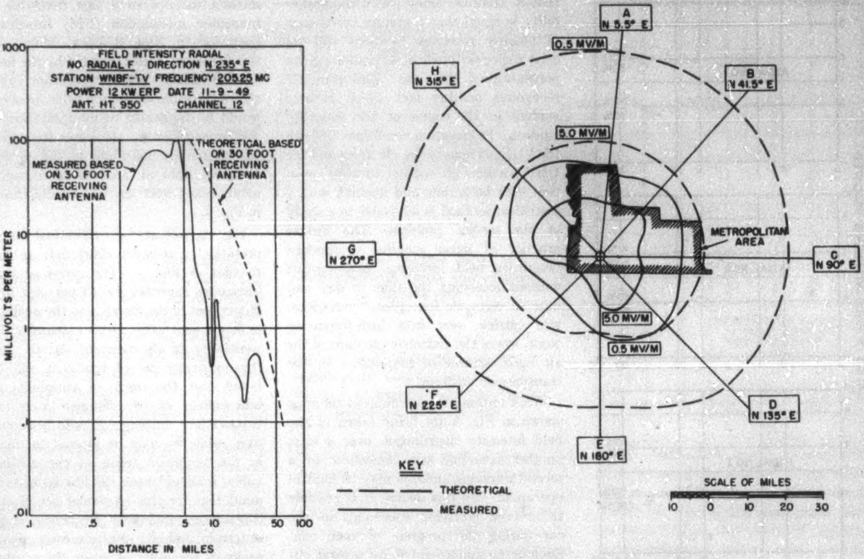


Fig. 10. Theoretical versus measured contours for a television broadcast station and analysis of radial F

Figs. 9 and 10 indicate typical data presented to the Federal Communication Commission in the form of proof-of-performance studies. The graphs and maps shown are more in the form of conclusions than of basic data, in that the curves are derived from sector analysis of mobile field intensity recordings previously described. Fig. 9 contains information with respect to a channel No. 2 (54 to 60-mc) operation in the Detroit Mich., area. Fig. 10 presents data for a channel No. 12 (204 to 210 mc) operation in the Binghamton, N. Y., area. The frequencies in these cases are illustrative of practically the total frequency range in the VHF portion of the spectrum. But one cannot draw comparisons here on the basis of two cases looking toward a frequency effect factor, because the terrain and even the types of cities involved are widely different. However, it may be appreciated that with proof of performance data for about 24 TV and FM operations available, some comparisons and trends could be developed. It was on the basis of just such data and methods of analysis, here very briefly outlined, that the working committees of the 1948–1952 period were able to piece together a hypothesis.

During the course of deliberations, which began in June 1948, it was apparent that tropospheric interference, which had been experienced in the FM band in the 100-mc region, could not be ignored in developing a nation-wide TV allocation plan. However, it was also apparent that very few basic data which could be used in studying the tropospheric effect existed, and that what data were available would require a great deal of study and analysis. Realizing that it would be unsafe to develop an allocation plan which ignored these effects, certain additional studies were proposed. It can be appreciated that under all these circumstances serious questions among engineers arose both as to the methods used in analyzing the meager data available and as to the results obtained. The imperative requirement of the Federal Communications Commission for an allocation plan had to be subordinated while new explorations of the transmission medium were made. The report of the working committee finally issued in 1949 made it clear that the results could best be classed as only rough approximations, although the curves contained therein have been used as the basis for the allocation. The following quotation from the introduction to their report is of interest: "The committee feels that the information it has had to work with is far too scant for it to consider the attached curves to be more than a relatively rough approximation to the propagation conditions to be expected throughout the United States."[2] It is important to emphasize also the fact that these deliberations were carried on without regard for UHF considerations where almost no experimental data were available at that time. The ultimate partial application to UHF problems of the curves developed in this report can therefore be regarded only as some order of further approximation based on the original rough approximation.

The aforementioned report[2] is limited in scope to the determination of engineering factors necessary to the prediction of service and interference from VHF stations. No attempt was made in it to determine how far apart stations should be placed on the same and adjacent channels, because it was felt that this involved policy decisions which would have to be made by the Commission itself. As noted, however, the report does contain the basic and fundamental engineering considerations necessary to the architecture of the TV allocation plan, particularly in so far as transmission medium factors of importance in the VHF portion of the spectrum are concerned. That these fundamentals were carried forward into UHF problems was not necessarily to be expected. Experience with UHF operation, in spite of the fact that no field intensity measurements are now required in TV proof-of-performance tests, may in time result in new yardsticks by which the wisdom of assumption, in part, of VHF factors for UHF application will be judged. If, as new assignments are activated, the allocation pattern appears to be substantially less than perfect and subject to the complications of the type encountered in the immediate postwar period, then we may again enter upon a new period of re-examination and re-evaluation.

The stated objectives of the Ad Hoc Committee provide in outline form a summarization of their approach to the problem.

> "A. Prediction of Service Field Intensities—This includes the effects of fine and gross terrain features upon the median signal levels and a realistic appraisal of the apparent antenna height of the transmitter.
>
> B. Evaluation of the Random Variations in Field Intensity from Median Levels due to Local Terrain and Buildings.
>
> C. Tropospheric Propagation Curves—The determination of methods for the evaluation and treatment of available data and of the methods and basis for extension of the data to other frequencies and antenna heights.
>
> D. Method of Combining the Effects of the Spatial and Time Variations of the Desired Signal and One or More Interfering Signals."

It is beyond the scope of this paper to deal in detail with all the factors and the reasoning behind the work of this committee, and to treat their findings exhaustively. However, before presenting the standards with respect to urban and rural grades of service employed in television allocations work, we can point out a few of the salient elements in the considerations.

Ground wave signal range curves customarily used by the Federal Communications Commission are based upon an unattenuated field reference of twice free-space field values for smooth-earth theoretical field calculations. The committee decided that, for the purpose of making statistical comparisons between measured and calculated values of field intensity, the standard curves should be modified to follow the free-space inverse distance value, rather than twice this value, to the point of departure of the smooth-earth theoretical curve below that value. This revision was made since the line corresponding to twice inverse distance values represents the envelope of values expected under smooth-earth conditions, rather than the median value as exemplified by the curves developed.

Although a number of methods were considered for expressing transmitting antenna height, the postwar practice of specifying the average height above terrain between 2 and 10 miles from the transmitter was retained. The Committee pointed out that, while alternative methods could be used and might indeed appear to be more applicable in special cases, if other methods were employed the statistical presentation relied upon in the instant report would not be valid for an alternative method.

A large portion of the data studied by the committee consisted of mobile recordings taken at receiving antenna heights of from 8 to 12 feet above ground. To relate these data to a 30-foot receiving antenna height used as a par by the Commission for several years, it was finally decided to use a linear height-gain relationship, although there appeared to be considerable doubt as to the

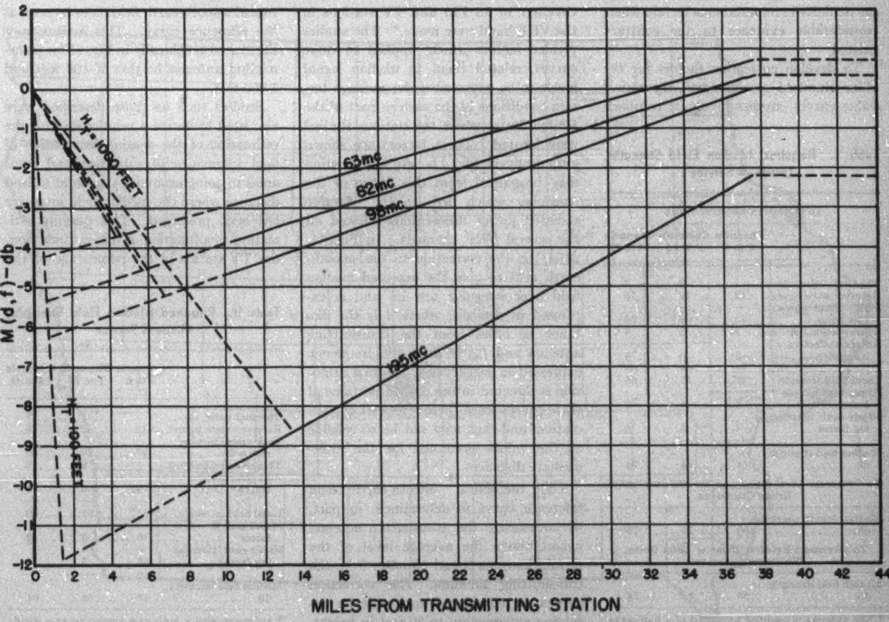


Fig. 11. Correction to the theoretical smooth-earth field intensity required to give the median of the fields to be expected over irregular terrain

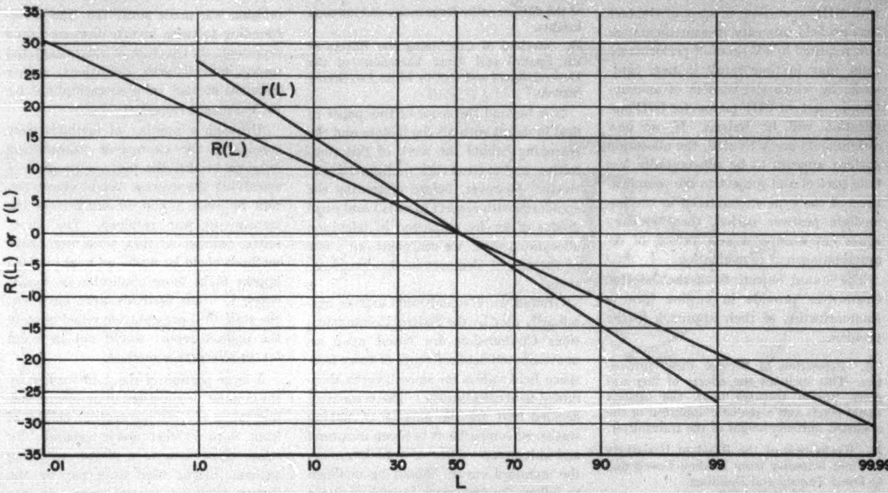


Fig. 12. Ratio in db of the field intensity exceeded at L per cent of the receiving locations to the field intensity exceeded at 50 per cent of the receiving locations for VHF and UHF TV frequencies

actual realization of such a relationship, particularly in mountainous terrain where considerable evidence to the contrary existed.

To develop correction factors for the older standard smooth-earth theoretical value curves, studies involving comparisons of the predicted and achieved coverage of 13 FM and TV stations in the VHF band were made.[3] The studies yielded factors which, applied to these curves, related them to median signal levels to be expected under average terrain conditions in the eastern part of the United States where the stations studied were located. These factors are shown graphically in Fig. 11. An appreciation may be gained from this figure of the reasoning which led to the effective radiated power differentials allowed on the several VHF channels. In Fig. 11, $M(d,f)$ is the correction to the smooth-earth field to give the expected median field over irregular terrain, and is expressed in decibels, where d is the distance in miles from the transmitting antenna and f_{mc} is the radio frequency expressed in megacycles. Special attention is directed to the dashed portions of these curves since these represent extrapolations and thus may not be so reliable as the values predicted for the intermediate distances.

Table I. Required Median Field Strengths, Grade-A Service

To Overcome Receiver Noise	Channels 2 to 6	Channels 7 to 13	Channels 14 to 83
Thermal noise, db	7	7	7
Receiver noise figure	12	12	15
Peak visual carrier/ rms noise	30	30	30
Transmission line loss	1	2	5
Antenna effective length*	−3	6	8
Local field strength	47	57	65
70-per-cent terrain factor	4	4	6
90-per-cent time-fading factor	3	3	3
Median field strength, db	54	64	74
To Overcome Local Noise and Interference under Urban Conditions			
Median field strengths, db	68	71	74
To Overcome Receiver Noise or Local Noise, Whichever Is Greater			
Median field strengths, db	68	71	74

* The antenna is assumed to consist of a half-wave dipole for channels 2 to 13 and an antenna with 8-db gain for channels 14 to 83.

The theoretical smooth-earth field reference curve is determined, in part, by measuring the transmitter antenna height above the average level of the terrain between 2 and 10 miles from the transmitting antenna. The correction curves are considered applicable for receiving antennas up to 30 feet in height. However, at distances less than $12h_t'h_r'/\lambda$ the free-space field, rather than the theoretical smooth-earth field, is to be used as the reference curve. This limitation is illustrated graphically in Fig. 11 for transmitting antenna heights of 100 feet and 1,000 feet.

Studies such as those described were also used to derive a relationship for the estimation of the spatial distribution of field intensities in suburban and rural areas in going around a station at a fixed distance where the median field intensity has been predicted. This permits estimation as a function of the distance from the TV station of the percentage of the

Table II. Required Median Field Strengths, Grade-B Service

	Channels 2 to 6	Channels 7 to 13	Channels 14 to 83
Thermal noise, db	7	7	7
Receiver noise figure	12	12	15
Peak visual carrier/ rms noise	30	30	30
Transmission line loss	1	2	5
Antenna effective length*	−9	0	3
Local field intensity	41	51	60
50-per-cent terrain factor	0	0	0
90-per-cent time-fading factor	6	5	4
Median field intensity, db	47	56	64

* Antennas with a 6-db gain compared to a dipole are assumed for channels 2 to 13, and a gain of 13 db is assumed for channels 14 to 83.

receiving locations at which a particular field intensity is available. Thus the median field for 50 per cent of the locations 50 per cent of the time may be converted to a field for L per cent of the locations by the application of a proper factor. Fig. 12 shows the relationship developed as a result of the statistical studies. In this figure $R(L)$ is shown as a function of L. As an example, Fig. 12 indicates that the field at $L=10$ per cent of the receiving locations, $F(10,50)$, may be expected to be 11 decibels (db) larger than the median field $F(50,50)$. For the available data and the analysis employed, $R(L)$ was found to be approximately independent of the distance, the frequency, and the transmitting or receiving antenna height. The function of $r(L)$ shown on Fig. 12 represents an assumed distribution factor for UHF frequencies.[4]

In its TV standards the Commission has provided for two grades of service. Grade A service is so specified that a quality acceptable to the median observer is expected to prevail for at least 90 per cent of the time at the best 70 per cent of receiver locations at the outer limits of this service. For Grade B service the comparable figures are 90 per cent of the time at 50 per cent of the locations. The required median field strengths in db above one microvolt for these grades of service on the several channels were developed as shown in Tables I and II.

The use of these grades of service values for the several channels in conjunction with the $F(50,50)$ median field intensity charts in the Commission's standards results in establishment of the theoretical radial distances to various points on the grades of service limits when the effective radiated power and the transmitting height above average terrain are known.

The opportunity to develop educational television broadcast systems on an integrated basis came when the Federal Communications Commission provided in its allocation plan 242 channels for noncommercial educational television use. Several of the states have been interested in the utilization of channels assigned within their state, and the application of the foregoing knowledge to a typical example has been chosen as illustrative of these considerations. Furthermore, because it has not been possible to treat the Commission's standards in the detailed way desirable in one paper on the subject, the illustration will serve to point up several other pertinent features of the standards.

The table of assignments contained in the present list of rules[5] indicates the following reservations for educational use in the state of Iowa: Cedar Rapids, channel 26; Davenport, channel 30; Des Moines, channel 11; Iowa City, channel 12; Sioux City, channel 30; Waterloo, channel 22.

While these assignments are fairly well distributed geographically, it would not be feasible to provide state-wide coverage if a system were to be limited to the use of these six channels. Fortunately, the rules of the Commission provide a process whereby additional channels may be assigned to certain cities, providing the separation requirements of the standards for the various channels can be met. Accordingly, in our planning, studies were carried out looking toward the addition of several channels to the basic ones assigned. It will be noted that two of the reservations listed were for VHF channels, i.e., Des Moines and Iowa City, and the remainder are for UHF channels.

Prominent among the considerations in developing a state-wide system are the population characteristics of the region

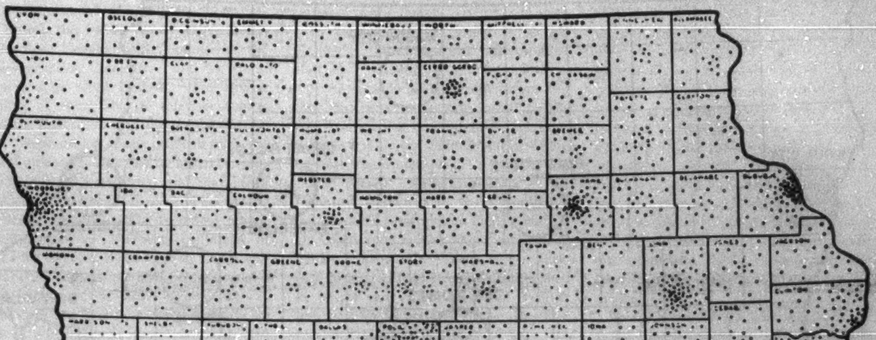


Fig. 13. Distribution of Iowa population. Each dot represents 1,000 persons. Population 2,643,000, Census Bureau estimate, July, 1949

Table III. Population of Iowa Cities Over 20,000

Des Moines	177,965
Sioux City	83,991
Davenport	74,549
Cedar Rapids	72,296
Waterloo	65,198
Dubuque	49,671
Council Bluffs	45,429
Ottumwa	33,631
Burlington	30,613
Clinton	30,379
Mason City	27,980
Iowa City	27,212
Ft. Dodge	25,115
Ames	22,898

to be served. The population of the state of Iowa, according to the 1950 census, totals 2,621,073. The state has a land area of 56,045 square miles, so that on the average there are approximately 47 inhabitants per square mile. Among the states, Iowa ranks twenty-second in both land area and population.

The urban population of Iowa numbers 1,250,938 persons or 47.7 per cent of the population of the state. In 1950, the urban population resided in 93 incorporated places and in other territory in the urban fringes of the three urbanized areas entirely in the state, and of the Iowa parts of three other urbanized areas partly in the state.

The rural population of Iowa comprises, 1,370,135 persons or 52.3 per cent of the population of the state. Of the rural population, approximately 200,000 reside in 126 incorporated places of 1,000 to 2,500 inhabitants.

There are 14 cities in the state having a population in excess of 20,000 and these are listed in Table III in the order of their size in descending order of magnitude. It should be noted that if Cedar Falls (population 14,334) and Waterloo were taken together as one unit, their population would total 79,532, which would rank this combination third among the cities listed.

Reference to the population dot map, Fig. 13, shows graphically the population concentrations listed in Table III. The 1950 report of the Bureau of the Census defines an urbanized area as an area containing at least one city with 50,000 inhabitants or more in 1940. An urbanized area may also contain two or more cities of 50,000. There are three urbanized areas entirely within Iowa and three partly within the state. There were 362,532 persons living in the first three areas, the Cedar Rapids, Des Moines, and Waterloo urbanized areas. There were 212,969 persons in the Iowa parts of the Davenport (Iowa) Rock Island, Moline (Illinois), Omaha (Nebraska), and Sioux City urbanized areas. Of the 575,501 persons in the state living in urbanized areas, 473,999 were in the five central cities, Cedar Rapids, Davenport, Des Moines, Sioux City, and Waterloo, and 101,502 were in the urban fringes.

Table IV. Limits for Field Intensity

Channels	Grade A	Grade B
2 to 6	68 dbu	47 dbu
7 to 13	71 dbu	56 dbu
14 to 83	74 dbu	64 dbu

Also basic to the development of the system, of course, was the existing distribution of educational facilities within the state. It has been indicated in a report that provision should be made for establishing four major points for program origination.[6] These points were given as Ames, Cedar Falls, Des Moines, and Iowa City.

The availability of program material at each of these localities would provide a cross section of educational program material which might be used throughout the state. The report indicated that Iowa State Teachers College would fur-

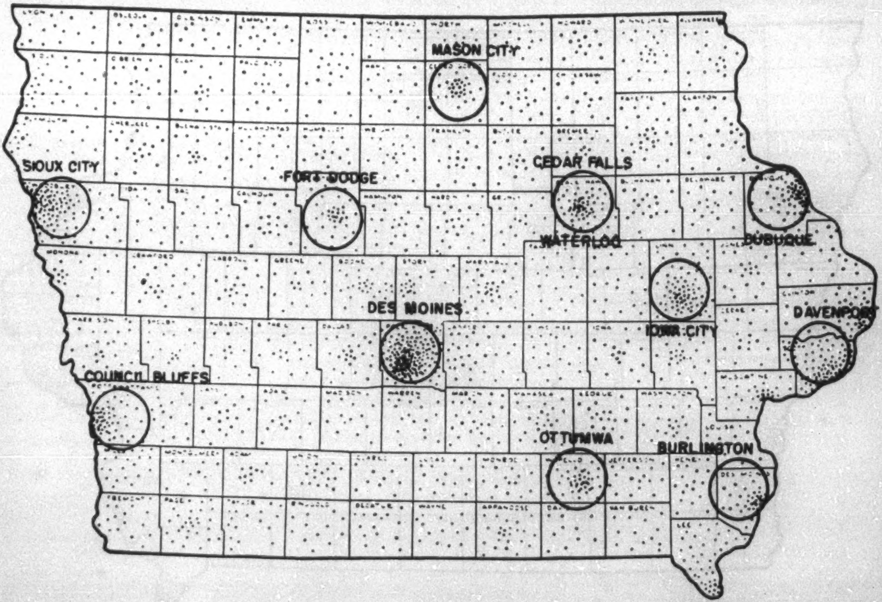


Fig. 14. Idealized distribution of educational TV broadcast stations to serve state of Iowa. Each dot represents 1,000 persons. Population 2,643,000, Census Bureau estimate, July, 1949 (circles do not indicate coverage)

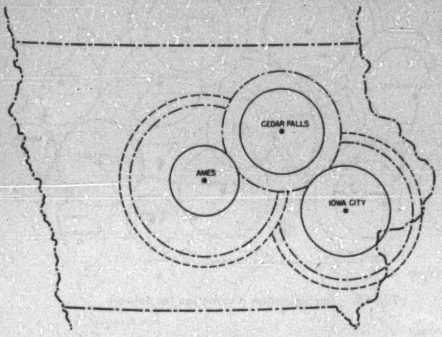


Fig. 15. Basic 3-station distribution for network, phase 1

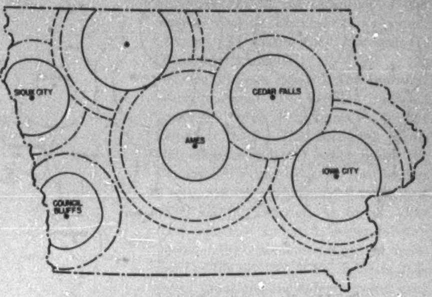


Fig. 16. Six-station distribution for network, phase 2

nish much of the material used for in-school purposes, while the State University of Iowa would contribute programs dealing with liberal arts and medicine, and Iowa State College would furnish programs relating to agricultural subjects, home economics, and engineering. Programming at Des Moines could include not only programs developed by the State Department of Public Instruction but also the many programs available from other state departments such as the Department of Health, the Board of Control, the Public Safety Department, the Iowa Development Commission, the Conservation Committee, and the Commission on Civil Defense. The desirability of being able to televise sessions of the legislature was also mentioned.

Consideration of the problem from the standpoint of an idealized station distribution is interesting, and involves a brief recapitulation of some of the points already covered. The technical standards regarding television contained in the list of rules of the Federal Communications Commission consider two grades of field intensity limit. These are defined as Grade *A* and Grade *B* limits for field intensities as noted in Table IV. The decibel-above-1-microvolt (dbu) reference level is 1 microvolt per meter.

The technical standards further provide, in connection with the location of the transmitter and antenna system, that the minimum field required over the entire principal community to be served shall be 74 dbu on channels 2 to 6, 77 dbu on channels 7 to 13, and 80 dbu on channels 14 to 83. It will be noted that these values are 6 db greater than grade-*A* values.

The requirements for the grades of service outlined were determined by making certain assumptions regarding receiver noise, transmission line losses antenna efficiencies, local noise and interference conditions, terrain factors, and time-fading factors. Based on these assumptions, the standards indicate that grade-*A* service for the VHF and UHF channels is, in effect, determined by the local noise and interference criteria under urban conditions, since the signal requirement based upon this condition equals or exceeds the signal levels required to meet other reception factors outlined in the foregoing.

In establishing the grade *B* service standards, the Commission contemplated that this reference would be operative only for rural location considerations and that the receiver factors, including the transmission line and antenna efficiencies and the terrain and fading factors, would be solely determinative.

The present charts contained in the technical standards of the Commission for predicting estimated field strengths are based upon median signal expectancy values. Specifically, these charts indicate estimated field strengths established at 50 per cent of the potential receiver locations for 50 per cent of the time at a receiving antenna height of 30 feet. To obviate the need for many charts showing field intensity for different percentages of time and location, all calculations are based upon the median value curves. Elevating the median values by the proper number of db secures the desired percentages contained in the standards for the various grades of service. Therefore, when performed in accordance with the present Commission technical standards, the interpretation of TV service is a statistical matter. In effect, the grade-*A* locus then becomes a limit at which, based upon average conditions, the service expectancy is for 70 per cent of the receiving locations 90 per cent of the time. Similarly, the grade *B*-limit line defines the locus at which the service epectancy is for 50 per cent of the locations 90 per cent of the time. It may be noted that the 90-per-cent time factor is a constant throughout and that on a statistical basis the percentage of locations at which satisfactory signals would be received would increase as the distance from the transmitter is decreased from either of the limits.

Based upon these considerations, it may be realized that any plan for statewide coverage of Iowa should endeavor to provide grade-*A* or better service over the urban areas of the state. In general, it may be stated that the more populous the area, the more stringent the service requirement. This leads to a recommendation for location of stations centered on the localities of urban concentration. Certain practical compromises with this approach are necessary, but for reference purposes, Fig. 14 portrays an idealized station distribution on the dot population map of Iowa. Each dot represents 1,000 persons and the location of the major urban areas is very evident.

In the case of UHF station considerations, the assumptions of the Federal Communications Commission for the various grades of service are numerically equal to the total of the receiver factor limitations under urban conditions without regard to noise and interference considerations. While receiver characteristics for the UHF frequencies may be expected to improve somewhat in the future, present factors indicate an even greater need for placing UHF facilities in the heart of the area to be served, whereas, because of superior receiver characteristics on the VHF channels, considerably more leeway in location is possible.

From the operation of TV stations to

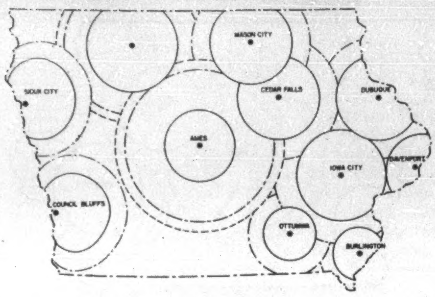


Fig. 17. Eleven-station distribution for network

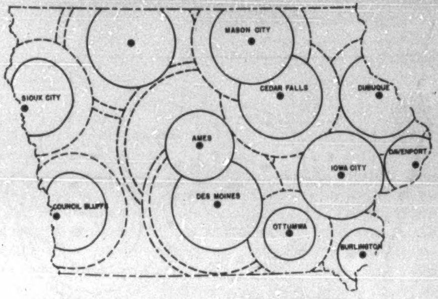


Fig. 18. Twelve-station distribution for network

date, we are well aware that the receiving public will, in many cases, make better than average installations in the way of antenna heights and antenna gain than those assumed in the development of the grade-*A* and grade-*B* service definitions outlined. Our experience is, of course, mainly limited to VHF considerations, since to date only a few of the TV stations operating are on UHF channels. In showing VHF service to the state of Iowa, we have assumed that 6-db more gain could be realized by the installation of better than average VHF receiving antennas or that a 12-db factor of gain is not unreasonable for a VHF antenna installation at a considerable distance from the transmitting station. It has not been possible to make a similar assumption to postulate a grade of service beyond the grade *B* limit in the case of UHF channels, because this service definition of the Commission already includes an assumption that a 13-db gain antenna will be employed on channels 14–83.

Experiments to date relative to the advantages to be gained by placing the receiving antennas at greater heights have been rather inconclusive. Therefore, while we recognize that many persons remote from the transmitter employ antennas installed at heights greater than 30 feet above ground, no extension of service factor based on this consideration has been incorporated in our hypothesis. It is possible, of course, that further experience with UHF receiving conditions will enable us to postulate improved coverage for this type of station in the future, based upon a combination of this and other factors. We do not, therefore, regard the limits indicated on the coverage maps as definitive in the sense that there will be a complete absence of signal in the several zones indicated. As a practical matter, we have endeavored to make these zones fall in areas where the population is most sparse, and where the local interfering factors are minimized.

The subcommittee[6] report has proposed that, as a first step in an over-all plan, TV broadcasting stations and program-originating facilities be established at Cedar Falls and Iowa City. These two additional stations would then be linked by program relay facilities with the existing station, WOI-TV, at Ames, Iowa, to provide a basic 3-station network. The report also investigated the establishment of a program-originating facility at Des Moines, Iowa. This latter facility could also be interconnected with the other three program-originating centers to form the nucleus of a state-wide system.

In the TV broadcasting service, as in other broadcast services, it is fairly common practice to have separate program-originating or studio facilities and transmitting plant facilities. As in the case of the present operation at WOI-TV it may ultimately be more expedient in securing optimum placement of transmitting facilities from a coverage standpoint to locate such transmitters remotely from the studio location, which may be kept at a more convenient and centralized location. This is done to secure the advantages of superior terrain elevation or to provide an improved geographical distribution for the transmitted signal. Interconnection of the studio and the transmitter in such cases is usually by means of a microwave radio relay link.

Fig. 15 indicates the expected distribution of TV broadcast coverage from the basic 3-station operation. This figure also forms the basis for consideration of a second step toward a state-wide system. It would appear logical at this point that provision should be made for service to the western portions of Iowa.

Reference to the population concentration map shows that the most stringent requirements in this area will be imposed by the population concentrations at Sioux City and Council Bluffs. Since the Commission Assignment Plan for Educational Television in Iowa provided channel 30 for Sioux City, but none for the Council Bluffs region, a study was made to determine if a channel would be available for use in this area upon the presentation of a proper rule-making request to the Commission. Study indicated that channel 55 could be employed in this area. This and similar other studies indicated that the coverage distribution for the combined first and second steps

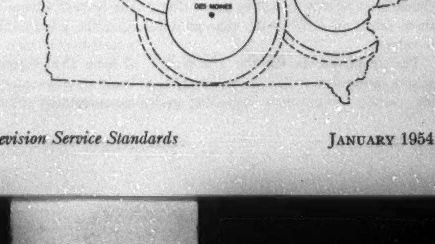


Fig. 19. Optimum distribution of stations for a VHF network

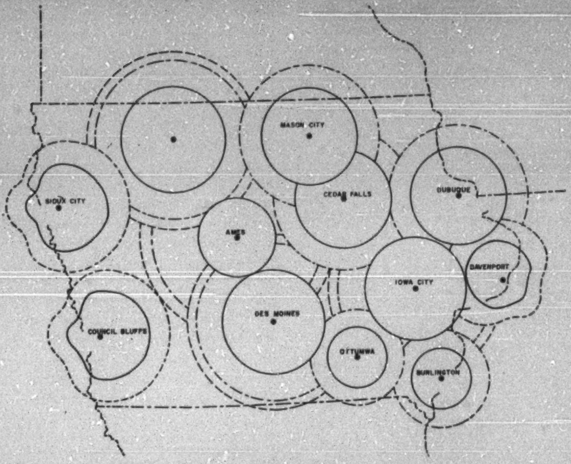


Fig. 20. Optimum distribution for 12-station network

in the envisioned program could be as shown in Fig. 16.

Reference to Fig. 16 and the dot population map of the state, Fig. 14, indicates the important urban and rural areas remaining in the state for which grade-*A* and grade-*B* service provisions should be made. By providing for a grade-*A* or better, service in Mason City, Dubuque Davenport, Burlington, and Ottumwa, most of the urban and rural gaps in Fig. 16 can be filled. Fig. 17 indicates the possibilities along this line.

Reference to Figures 17 and 14 indicates the possibility and the desirability of employing the channel 11 assignment at Des Moines to provide an improved grade of service in the major population center of Des Moines and in the rural central southern region of the state. Fig. 18 which has been prepared to illustrate this additional step in the development, is illustrative of one further important condition which involves consideration of the optimum distribution of the VHF channel assignments available for educational use in Iowa.

Since the receiving sets in the hands of the Iowa public today are 12-channel sets equipped to receive only VHF channels, we regard the VHF assignments available for the state program as a keystone of the state-wide service system for several years to come. Accordingly, it is important to consider as a separate facet of the problem the best employment of these facilities to secure the optimum coverage which they can provide within the limits of transmitter placement. The idealized distribution of these four VHF facilities is indicated in Fig. 19. Basically, this involves employment of the channel 11 facilities at Des Moines as far east and south of the city as it is practical to go, from a city service requirement standpoint, and the shift of the Ames transmitting facility to a locality more central with respect to the Des Moines and the assignment in the northwestern part of the state. This alternate arrangement would also more nearly satisfy the service requirement of Fort Dodge for a grade-*A* or better service. The composite coverage based upon activation of all the steps outlined is shown in Fig. 20.

The additional channels for the other cities requiring a grade-*A* service are summarized below. These are all UHF channels, since the VHF saturation technique employed by the Federal Communications Commission in developing its Table of Assignments precluded VHF assignments with the exception of one case: Burlington, Channel 63; Dubuque, Channel 68; Mason City, Channel 47; Ottumwa, Channel 41.

Table V gives a list of the proposed operations and the effective radiated power and type of antenna to be employed by each. The maximum allowable effective radiated power under the Commission's rules is 100 kw for channels 2 to 6, 316 kw for channels 7 to 13, and 1,000 kw for channels 14 to 83. The term effective radiated power as employed in the Commission standards is defined as the product of the antenna peak power input and the antenna power gain. The horizontal plane is the reference dimension for the antenna gain term. The conventional TV antenna employs elements, so phased and so arranged along a supporting member that radiation is concentrated toward the horizon, and radiation at the vertical angles minimized.

Effective antenna heights of 500 feet would be employed in all cases. Effective antenna height is defined by the Federal Communications Commission as the height of the electrical center of the TV radiating system above average ground elevation, determined in eight radial directions over a 2-to-10 mile interval of distance from the transmitter site. Supporting tower height requirements will therefore vary somewhat, depending on ability to locate transmitting tower sites at higher-than-average ground elevations for the several locations.

Directional antennas have been assumed in three instances to minimize the out-of-state coverage and the required number of stations. The rules of the Federal Communications Commission contemplate allowing directional antenna operation, providing that the degree of directionality employed is kept within certain limits, and that in no event is the maximum effective radiated power specified under the rules exceeded in any direction. Therefore, in the case of the Council Bluffs and Sioux City operations while the effective radiated power in an easterly direction is 1,000 kw, the average

Table V. Power and Antennas for Proposed Operations

City	Channel	Effective Radiated Power	Type of Operation
Ames	5 (VHF)	100 kw	nondirectional
Burlington	63	100 kw	nondirectional
Cedar Falls	22	1,000 kw	nondirectional
Council Bluffs	55	525 kw	directional
Davenport	30	200 kw	directional
Des Moines	11 (VHF)	1,000 kw	nondirectional
Dubuque	68	1,000 kw	nondirectional
	12 (VHF)	316 kw	nondirectional
Iowa City	12 (VHF)	316 kw	nondirectional
Mason City	47	1,000 kw	nondirectional
Ottumwa	41	100 kw	nondirectional
Sioux City	30	525 kw	directional

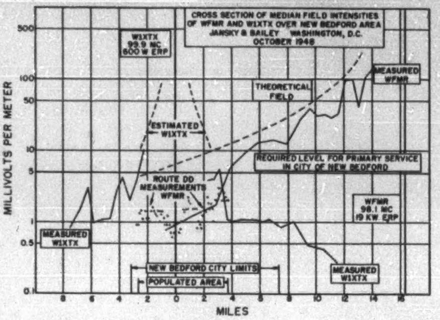


Fig. 21. Theoretical and measured field intensity distribution over a city illustrating importance of transmitter siting

power for the whole pattern employed is considerably less than this value. Maximum allowable power was proposed in all cases to provide the maximum possible signal in all areas, consistent with Commission standards.

In the case of the four VHF stations, commercial equipment to attain the maximum effective radiated powers specified is available on the present market. As regards UHF operations, however, commercially available equipment at present imposes a power ceiling of approximately 275 kw of effective radiated power. Higher powered equipment is still in the design stage at present, although certain preliminary cost estimates for equipment combinations to produce 1,000 kw of effective radiated power are available. Therefore, of necessity, an initial operation at Cedar Falls, under phase 1 of the program might be for a lower effective radiated power than that assumed. The prompt establishment of this station, however, albeit at a lower power, would enable measurements to be made and experience to be gained which could furnish an important guide in the development of the other phases of the system.

The importance of optimum transmitter siting in the development of a program such as outlined cannot be overemphasized. As an example, while it would be tempting to say that one transmitter located centrally with respect to Iowa City, Davenport, Burlington, and Ottumwa could adequately serve these four cities, the percentage of locations in each city where reception difficulty prevailed would undoubtedly be high. Fig. 21 illustrates a case in which a 600-watt FM station centrally located in the city concerned was found to be necessary to supplement the service from a 19-kw FM station located approximately 16 miles from the center of the city. While on a theoretical basis adequate coverage of the city was indicated, this did not prove to be the case in practice, as is shown. The path for route DD measurements, which was shown on Fig. 7, lies generally along the last portion of the path of the main radial run from the WFMR transmitter site. This case is an example of the importance of paying due respect to the statistical theory of city coverage. This theory may be stated approximately as follows: In any city in which the field intensity of a broadcast station is not at every point sufficient to provide a satisfactory service in the presence of the noise level at that point, there will be some percentage of the potential listeners who will not receive satisfactory service. This percentage will vary with the average field intensity, with the distribution and intensity of noise sources, and with the variability of the signal over the city.

References

1. ON THE USE OF FIELD INTENSITY MEASUREMENTS FOR THE DETERMINATION OF BROADCAST STATION COVERAGE, **C. M. Jansky, Jr.**, **Stuart L. Bailey.** *Proceedings*, Institute of Radio Engineers, New York, N. Y., vol. 20, Jan. 1932, pp. 62–76.

2. REPORT OF THE AD HOC COMMITTEE FOR THE EVALUATION OF THE RADIO PROPAGATION FACTORS CONCERNING THE TELEVISION AND FREQUENCY MODULATION BROADCASTING SERVICES IN THE FREQUENCY RANGE BETWEEN 50 AND 250 MC. Federal Communications Commission, Washington, D. C., 1949.

3. **Central Radio Propagation Laboratory.** National Bureau of Standards, Washington, D. C.

4. PROPOSED VHF-UHF RULES STANDARDS AND ALLOCATIONS. *Report No. 3*, Federal Communications Commission, Washington, D. C., March 24, 1951.

5. RULES OF THE FEDERAL COMMUNICATIONS COMMISSION GOVERNING TELEVISION STATIONS. Federal Communications Commission, Washington, D. C.

6. REPORT OF THE EDUCATIONAL TELEVISION BROADCASTING SUB-COMMITTEE OF THE COMMITTEE ON INTER-INSTITUTIONAL COORDINATION OF THE STATE BOARD OF EDUCATION.

◆

No Discussion

Automatic Testing of Wired Relay Circuits

A. N. HANSON
MEMBER AIEE

PRACTICALLY all of the telephone switching equipment in an exchange is made up of wired assemblies—assemblies of relays and other components interconnected by a multiplicity of wires. These assemblies are on frameworks generally 11 feet 6 inches high. Many of these frameworks are made up of smaller components called relay circuit units composed mainly of relays. Rotary switches, resistors, capacitors, jacks, etc., necessary for the operation of switching circuits may also be included. These components vary in height up to about 24 inches and after testing they are mounted in the larger frameworks. A careful check of connections and electrical continuity is essential to assure that these units will function properly when connected to other circuits in a telephone exchange.

Buzzer continuity tests lack much in effectiveness and manually operated test sets are slow and tedious, and lead to human errors. Manual test sets, however, have the advantage of great flexibility in testing various types of circuits not usually found in automatic test sets. Automatic test sets have been in use for

---

Paper 53-407, recommended by the AIEE Communication Switching Systems Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Kansas City, Mo., November 2–6, 1953. Manuscript submitted August 11, 1953; made available for printing September 23, 1953.

A. N. HANSON is with the Western Electric Company, Inc., Chicago, Ill.

Grateful acknowledgement is accorded to A. J. Ehlschlager, designer of the test set, and to F. L. Osborne, for their aid in preparing this paper.

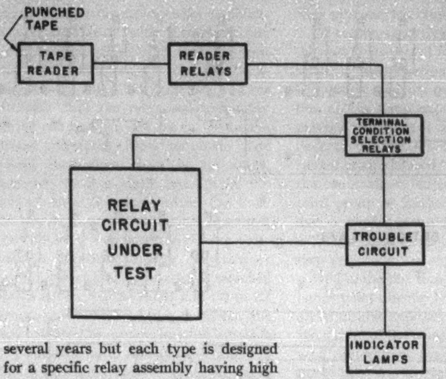


Fig. 1 (left). Test set functional block diagram

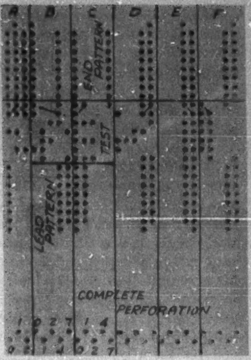


Fig. 2 (right). Section of perforated tape

several years but each type is designed for a specific relay assembly having high production requirements, and which is felt to be of stable character. The automatic sets are prepared for test by operating a few test set switches to a fixed position. A test set start-switch is then operated and the test proceeds to completion unless trouble is encountered. In the latter event, the test set stops and display lamps light to aid in locating the trouble. These automatic sets lack the flexibility of the manual sets and, to realize efficient production, they require the test man to be fully familiar with the circuit under test.

A new test set has been developed to eliminate the undesirable features of manual testing and still retain complete flexibility. The new test set reduces, sharply, the human equation as it automatically checks relay assemblies wired to perform practically any switching functions. Employing perforated tape for switching control, the test set applies operating and test conditions to terminals of the wired assembly. Indicator lamps pin-point terminals which fail to meet test conditions and, in most cases, indicate the nature of the trouble. This automatic test set is the subject of this paper. The particular set to be described was designed for testing relay circuits having up to 64 connecting block terminations. The design principles may be applied to testing larger assemblies up to practical limits of test set size and the switching principles could be used for testing other types of wired assemblies.

Table I. Available Terminal Testing Conditions

Condition No.	Reader Relays Operated	Terminal Relays Operated	Terminal Condition Desired
0*			Open check
1	*D*0, *D*1	*A*, *B*	Ground
2	*D*0, *D*2	*A*, *C*	200-ohm ground
3	*D*1, *D*2	C	Ground check
4	*D*0, *D*4	*B*, C	Battery
5	*D*1, *D*4	*A*, *B*, *C*	Common jack
6	*D*2, *D*4	*B*	Battery check
7	*D*0, *D*7	*A*	Individual jack

* Condition 0 is applied to any terminal not having conditions 1 to 7 and is not affected by the *D* group reader relays.

Test Sets, General

A brief discussion of the types of manually operated sets used to test wired relay assemblies will help to show the advantages of the automatic set. First, a manual set requires many cross-connections that vary for each different type of wired relay circuit to be tested. Since the circuits may be entirely different, one from another, each type of relay circuit tested has different circuit conditions wired to the connecting blocks. To bring the desired test set condition to the terminals of the wired assembly, cross-connections must be made at the test set that may number as many as 100. This is a time-consuming setup procedure and is a source of potential operator error. Second, practically every step in the progress of a test is controlled by manual operation of switches. Third, the operator verifies wiring accuracy by observing signal lamps which light or extinguish as each switch is operated. Fourth, the test man must refer to an instruction between each step of the test.

These are all major disadvantages as they all present potential sources of operator error. The automatic set avoids these disadvantages. With this automatic test set, progress of a test is controlled by relays which perform switching operations according to a sequence set by a perforated paper tape. The test continues automatically. When trouble is encountered, the set stops and indicates the location of the trouble on signal lamps. Because it is quickly adaptable to different wired relay assemblies, it has been called a Universal equipment test set.

Universal Equipment Test Set and Perforated Tape

Fig. 1 shows in block diagram form the main components of the Universal equipment test set. The tape is of specially prepared paper 3 inches wide, perforated with embossments in a saw-tooth fashion, with patterns of holes that can be read by the reading pins of an automatic message accounting reader. When placed in the tape reader, the embossments are on the underside of the tape. A selection of 28 embossments are possible across one line of tape. Fig. 2 shows a section of perforated tape demonstrating a complete line of twenty-eight holes and a

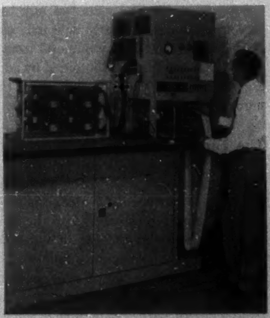


Fig. 3. The Universal equipment test set with unit under test

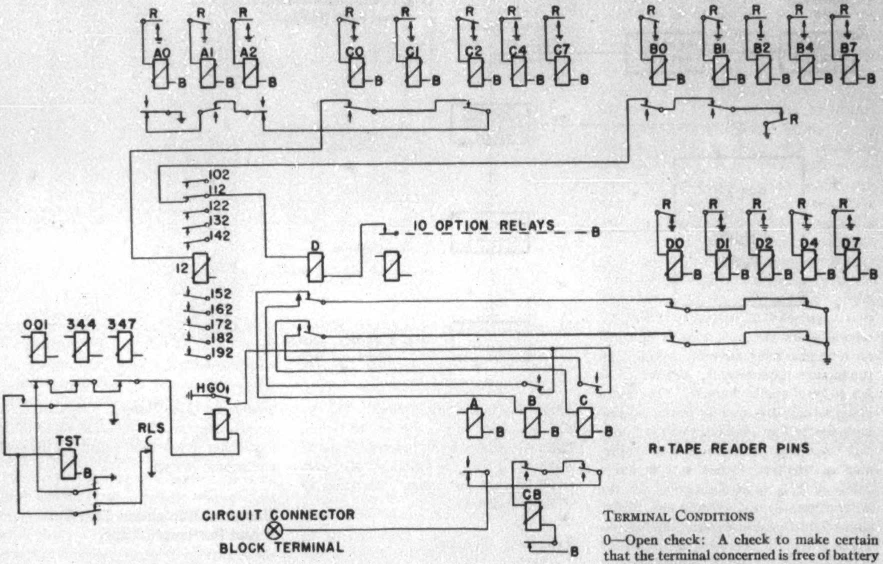


Fig. 4. Functional schematic of terminal selection and terminal condition selection circuit

compressed section of test perforations. The perforations, as shown in Fig. 2, are divided into six groups, *A* to *F*. Group *A* has three positions for the perforations and groups *B* to *F* have five positions each. Perforating a single position in the *A* group represents the numeral 0, 1 or 2. Perforating all positions in the *A* group represents the numeral 3. Perforations positions in groups *B* to *F* (there are 5 positions per group) are given values 0, 1, 2, 4 and 7 respectively. Two, and only two, perforations are provided in each of groups *B* to *F*. The values of any two perforated holes, properly selected, can be added to result in any numeral from 1 to 9. The numeral 0 is obtained by adding 4 and 7. The significance of the numerical values will be apparent later.

The Tape Reader

The tape reader steps the tape at a rate of 16 steps per second. Between each step, 28 sensing pins descend to the tape in the same saw-tooth pattern shown in Fig. 2. When a hole permits a pin to pass through the tape, a pair of twin contacts close, to provide ground potential to complete a circuit path. If no hole is found, the contacts remain open as the pin strikes the paper. Reader relays, so called since they operate directly from the reader-pin closed contacts, are provided, one for each perforation position. They are designated *A*0, *A*1, *A*2, *B*0, *B*1, *B*2, *B*4, *B*7, *C*0 to *C*7 and *D*0 to *D*7. The *E* and *F* digits require no reader relays. They are used only to light display lamps which connect directly to the reader-pin contacts.

Terminal Condition Selection Relays

Test set terminations are connected to the unit under test through fixtures that contact the unit connecting block terminals. To each of these terminations is connected a set of four relays designated *A*, *B*, *C* and *D*. By the proper pattern of holes in the *A*0 to *A*2, *B*0 to *B*7 and *C*0 to *C*7 tape groups, any desired terminal of the connecting block can be selected to receive the test condition prepared by its set of terminal condition selection relays *A*, *B* and *C*. By a pattern of holes in the *D*0-*D*7 tape group, which produce the condition number, relays *A*, *B* and *C* can be operated in any one of eight combinations to provide one of eight conditions required for testing. These eight conditions are shown in Table I. The following will clarify the use of the terminal conditions.

Terminal Conditions

0—Open check: A check to make certain that the terminal concerned is free of battery or ground potential.

1—Ground: Connects direct ground from the test set to a unit terminal.

2—Ground check: A check to determine that ground potential is received by the test set from the circuit under test.

3—200-ohm ground: Connects a 200-ohm resistance ground from the test set to a unit terminal.

4—Battery: Connects direct battery from the test set to a unit terminal.

5—Common jack: Connects unit terminals to a pin jack for patching in special conditions or to form a loop between two or more terminals by connecting them to the common pin jack simultaneously.

6—Battery check: A check to determine that a battery potential is received by the test set from the circuit under test.

7—Individual jack: Connects a circuit terminal to an individual pin jack assigned to each connecting block terminal, for patching in special conditions or to present an open circuit to the terminal if desired.

Trouble Circuit

The terminal conditions are used to set up, within the circuit to be tested, circuit conditions that will verify agreement of the circuit with manufacturing drawings. For instance, ground will be connected by means of condition 1 to a connecting block terminal of the relay unit which carries a wire to one terminal of a relay coil that has battery on the far end of the coil. Terminal condition 2 is connected to another connecting block terminal that connects to a relay spring-make combination of that relay with ground on one spring. Thus, a relay of the unit, has been energized, and means provided for checking that the relay has operated, and for checking that the relay coil and relay spring combination are connected to the proper connecting block terminal. The actual verification check is not made, however, until a check line, to be described later, is perforated on the tape. As the *A*, *B* and *C* relays are operated, they are locked up until the *D* relay operates to introduce another condition or until a release-all combination is punched in the *A*0 to *A*2, *B*0 to *B*7 and *C*0 to *C*7 groups on the tape. When the check line is introduced by perforating the numbers 0, 1 and 2 in the *A* tape group, 0 and 4 in the *B* tape group and 0 and 1 in the *C* tape group (no. 341), the test set hesitates for 2 reader cycles (2/16 seconds). In this time, the TBL (trouble) relay is introduced to determine if terminal condition 2 is satisfied. If ground does not come back from the test set over the proper terminal, the TBL relay operates. At the same time, condition 0 is checked on all connecting block terminals not affected by conditions 1 or 2. If potential is returned to the set from any terminal which should be an open circuit, the TBL relay operates and stops the test.

To every connecting block terminal of the unit is connected four lamps with the terminal identification stamped on the lamp caps. In one group, labeled "In Use *G*," a lamp showing the terminal number will light if ground is expected back to the test set, but is missing. A second group, labeled "In Use *B*" will act the same where battery is involved. A third group of lamps labeled "Open *G*" and a fourth group labeled "Open *B*" will light, respectively, if ground or battery is present on any terminal not involved in the test that is in effect. These lamps point directly to the terminals that are in trouble and they aid in determining the nature of the trouble. The check line tape perforations are introduced at frequent intervals to narrow the field to be searched when trouble is encountered.

Using the Test Set

Fig. 3 shows a view of a test set designed for testing a relay unit having a maximum of 64 connecting block terminals. The cabinet to the right contains the control keys, indicating lamps, rheostats, pin jacks for special conditions, and cable connecting sockets. Below this cabinet is seen the perforated tape reader. In the enclosed cabinet under the table top are mounted the switching relays, capacitors, and resistors. A relay unit is seen connected to the test set by fixtures and conducting cable.

To start the test, the perforated test tape is fed into the tape reader. The reader pins first find a pattern of perforations known as the lead pattern. The lead pattern insures that the tape is properly aligned and feeding correctly into the tape reader. From the lead pattern, the tape perforations lead directly into the test pattern. Referring to Fig. 4, assume that the test will start with the placement of 48-volt negative battery on terminal 11 of terminal strip group no. 2. The perforations in the tape *A* group determine the 10's digit of the terminal number, the tape *B* group determines the unit's digit of the terminal number and the tape *C* group determines the desired terminal group or the terminal strip number. Thus, to find the desired terminal, the tape *A* group will have a perforation in the *A*-1 position, the tape *B* group will be perforated in the *B*-0 and *B*-1 positions, and the tape *C* group will be perforated in the *C*-0 and *C*-2 positions. The *D* tape position at this point determines the terminal condition to be placed on the selected terminal, which in this case

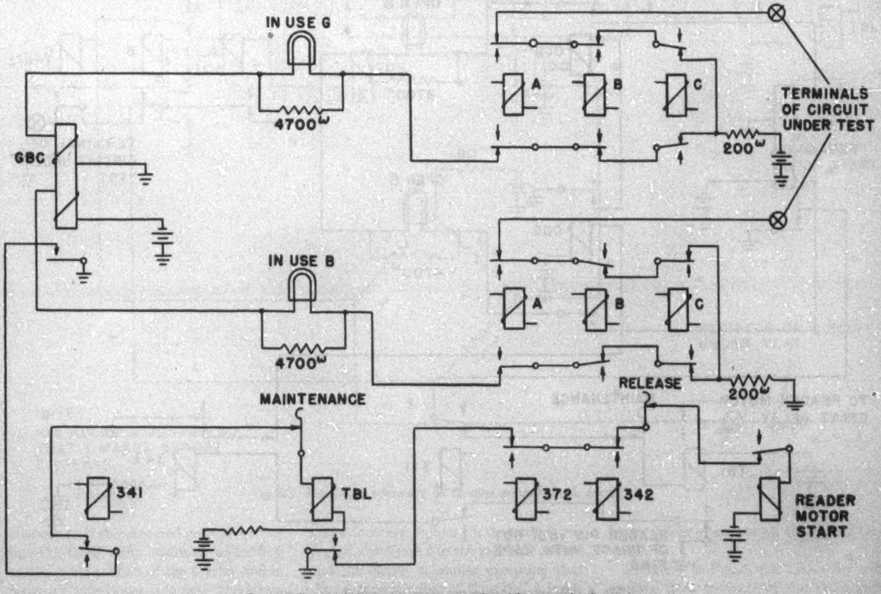


Fig. 5. Functional schematic of ground and battery check circuits

is condition 4 for −48-volt battery from the test set. The *D* group perforations, therefore, are *D*-0 and *D*-4. Thus, as the reader pins descend to the tape, they will find entrance through holes, registering the number 112,401 as the first line of the test pattern. The 01 represents the line identification and is not shown as a part of Fig. 4. Reader relays, corresponding in designation to the perforation designations, will operate directly from ground on the contacts of the reader pins that find perforations in the tape.

Referring to Fig. 4, the *A* and *C* tape groups select first the no. 12 relay representing the terminal 10's digit and the terminal strip number. Through one of the No. 12 relay make contacts, the tape *B* group selects the terminal identified by the units digit and operates the *D* relay associated with the terminal condition selection relays *A*, *B* and *C* associated with that terminal. The operation of the *D* relay closes a path to operate the *A*, *B* and *C* relays in a combination in accordance with the *D*0 and *D*4 perforations which in this case operates relays *B* and *C*, with *A* remaining normal. Operating relays *B* and *C* places battery on the relay circuit connecting block terminal. The *B* and *C* relays then lock and the *D* relay releases.

In similar manner, potentials are applied to other circuit-connecting block terminals to operate relays within the circuit under test. To check that the relays have operated, and to check that the wiring of the relay spring combinations are correctly connected, battery or ground potential is connected under tape control to the circuit terminals that are connected to the spring combinations. This potential from within the circuit is to be looked for at the test set when the check line is introduced on the tape. The check interval is introduced by presenting a tape line perforated in positions *A*0, *A*1 and *A*2, *B*0 and *B*4, and *C*0 and *C*1, or 341-XXX. At the check line, the *D*, *E* and *F* tape groups represent a test line identifying number. During this interval the TBL relay is connected into the circuit and a check is made by the test set to determine if the terminal conditions, set up by the terminal condition selection relays, are satisfied where potential is expected back from the relay circuit under test. At the same time, a check is made to determine that no potential is present on any terminal that should present an open circuit. Figs. 5 to 9 show typical examples of how the TBL relay is connected, to detect incorrect conditions.

Should the TBL relay remain normal, the conditions that the test set is looking for are satisfied and the test continues automatically. Should the TBL relay operate, it releases the reader clutch that advances the perforated tape and the test stops instantly. At that time, display lamps on the test set light to indicate the nature and location of the trouble. A set of lamps, giving the entire tape entry, light with every step of the tape, and if the tape stops in a check line this set of lamps indicates the serial number of the tape line of perforations. With each tape is provided a list of numbers that serve as a decoder for the perforated tape. Referring to this number list at the stopping or trouble point in the test, the test man can determine what test conditions were imposed after the last successful check line, and thus find aid in dealing with trouble.

Referring to Fig. 5, the TBL relay operates from contacts of an operated GBC (Ground battery check) relay, if either expected ground or expected battery from a predetermined connecting block terminal fails to appear. If ground is expected, on a circuit terminal, condition 3 on the test tape prepares the test set to receive it. If it appears, that ground will shunt out the ground winding of the GBC relay and prevent it from

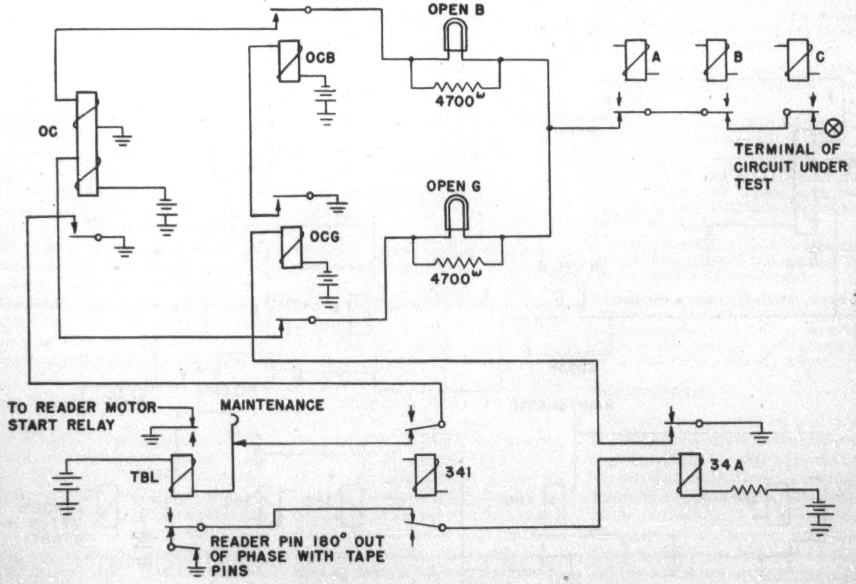

OPEN B
OCB
4700ω
A
B
C
OC
TERMINAL OF CIRCUIT UNDER TEST
OPEN G
OCG
4700ω
TO READER MOTOR START RELAY
MAINTENANCE
TBL
341
34A
READER PIN 180° OUT OF PHASE WITH TAPE PINS

Fig. 6. Functional schematic of open-circuit check circuits

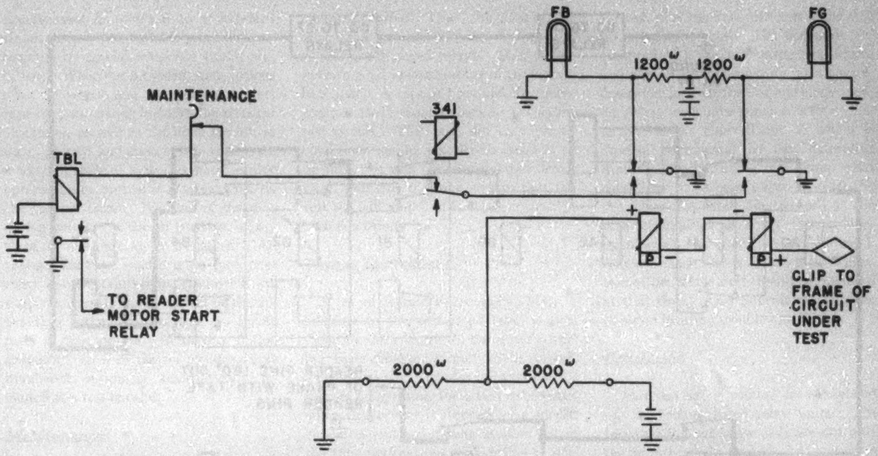

FB
FG
1200ω
1200ω
MAINTENANCE
341
TBL
+
−
P
P
−
+
CLIP TO FRAME OF CIRCUIT UNDER TEST
TO READER MOTOR START RELAY
2000ω
2000ω

Fig. 7. Functional schematic of frame ground and frame battery check circuits

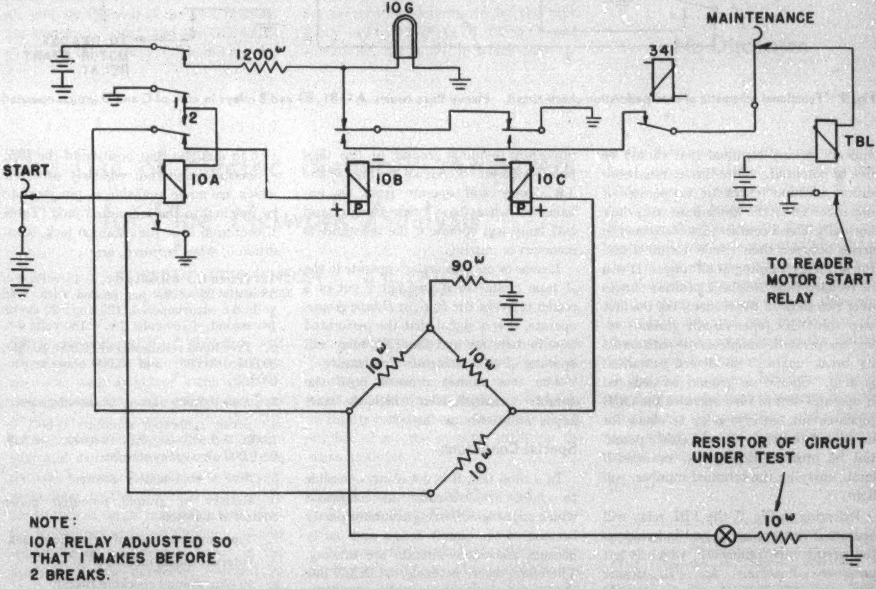

10G
1200ω
MAINTENANCE
341
1
2
START
10A
10B
10C
P
P
+
−
−
+
TBL
TO READER MOTOR START RELAY
90ω
10ω
10ω
10ω
RESISTOR OF CIRCUIT UNDER TEST
10ω
NOTE: 10A RELAY ADJUSTED SO THAT 1 MAKES BEFORE 2 BREAKS.

Fig. 8. Functional schematic of 10-ohm ground check circuit

operating. If the expected ground is missing, the GBC relay operates to battery on the make contact of the *C* relay and in series with the in-use-*G* lamp. This lamp, whose lamp cap is engraved with the terminal location, lights. If battery is expected, condition 6 in the test tape is used and the circuit is similar excepting that the battery winding of the GBC relay is involved and the in-use-*B* lamp lights if battery fails to appear on the circuit terminal.

Referring to Fig. 6, the TBL relay will operate from contacts of the OC (open check) relay if either battery or ground

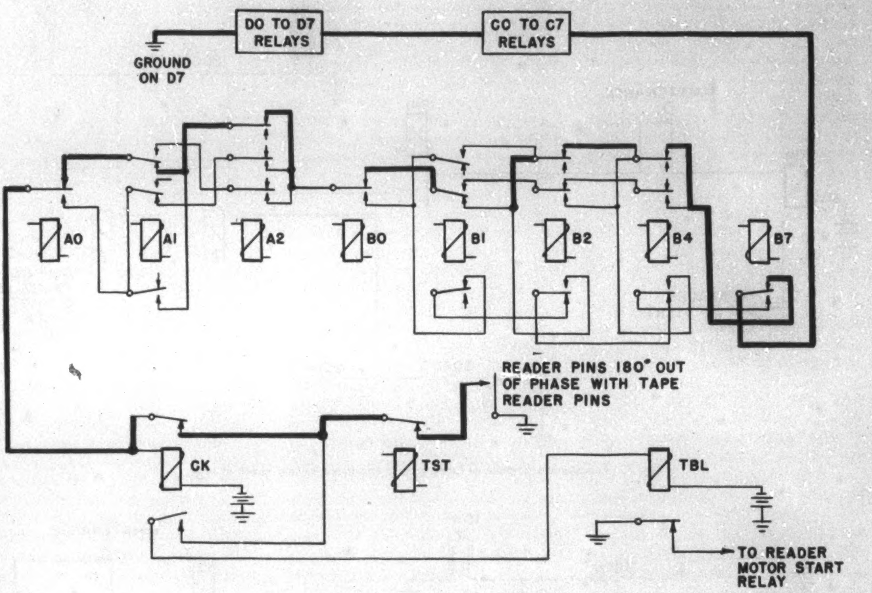

DO TO D7 RELAYS
CO TO C7 RELAYS
GROUND ON D7
A0
A1
A2
B0
B1
B2
B4
B7
READER PINS 180° OUT OF PHASE WITH TAPE READER PINS
CK
TST
TBL
TO READER MOTOR START RELAY

Fig. 9. Functional schematic of tape perforation check circuit. Heavy lines assume A1, B1, B7 and 2 relays in each of C and D groups operated

appears on any terminal that should be free of potential. Since the terminal condition selection relays are unoperated if not involved in the immediate test, their normally closed contacts are connected in series between their circuit terminal and the OC relay winding at all times. It will be recalled that the check position carries over two steps of the reader. On the first step the OCG (open-circuit ground) relay, unoperated, completes the path across its break contacts to detect unwanted ground. Should no ground be detected it operates and in turn operates the OCB (open-circuit battery) relay to check for unwanted battery. Should either potential be present, the open-*G* or open-*B* lamp, carrying the terminal number, will light.

Referring to Fig. 7, the TBL relay will operate if battery or ground is found on the circuit unit framework, which is left purposely floating. An FG (frame ground) or FB (frame battery) lamp will light in case of an error. Relays with the designation *P* on the relay core are of the polarized variety. That is, they are sensitive to the direction of the current through their windings.

Certain relay circuits are equipped with a grounded 10-ohm resistor used as a very important holding ground in the telephone system. Referring to Fig. 8, the TBL relay will operate from an unbalanced wheatstone bridge circuit and 10*G* lamp will operate if the resistance is incorrect or missing.

If none or only two relays operate in the *A* tape groups or if any but 2 out of 5 reader relays of the *B*, *C*, or *D* tape groups operate, it is a signal that the perforated tape is defective and the TBL relay will operate. Fig. 9 illustrates this feature.

The test cannot proceed until the trouble is cleared, after which the start key is used to resume the test.

Special Conditions

In a shop test, it is not always possible to check automatically all conditions which might affect wiring accuracy, partly because of the circuit design and partly because associated circuits are missing. Therefore, where necessary 342XXX line is inserted on the tape which stops the test set tape reader and lights the SP lamp on the test set panel. The last three digits of the 342XXX line will show the actual sequence number of the perforated tape line. The test set operating instruction, provided with every circuit, tells the test man what special verification to make at that position of the test.

Special circuits not regularly used but which are made available at pin jacks to be patched to the individual jack (Table I, condition 3) or the common jack, (condition 5) when required, are:

1. Ground, −24 volts d-c, −48 volts d-c, 88 volts 20 cycles per second with −48 volts d-c superimposed, 105 volts 20 cycles per second, +130 volts d-c, −130 volts d-c.
2. Five lump resistances of 10,000; 20,000; 30,000; 40,000; and 50,000 ohms respectively.
3. Two 0–9,999 ohm resistance decades.
4. Seven calibrated rheostats 0–500, 0–1,000, 0–2,500, 0–3,500, 0–5,000, 0–8,500, 0–10,000 ohms respectively.
5. Five 90-ohm battery circuits.
6. Battery or ground potential under control of a switch.
7. Tip, ring and sleeve of a telephone jack.
8. Interval timing circuit.
9. 10-ohm ground checking circuit.
10. 1,000 cycles per second source.
11. Transmission measuring circuit.

Options

An option consists of wiring or apparatus, or a combination of both in a circuit, used in addition to a standard circuit to provide variable conditions as required by special telephone traffic conditions. Wherever provided, such options must be tested and the perforated test tape for each circuit includes the tests for all options as well as the main circuit, in their proper sequence. Provisions are made to include on a tape as many as nine options under control of nine switches on the test set panel. With all of the nine option switches in the off position, all options on a tape may be by-passed. Although the tape reader must pass over every line of perforations, the switch will supply a circuit, under control of the appropriate entry on the tape, to ignore tests of that option. With any option switches 1 to 9 in the on position, that numbered option is automatically included as a test feature.

Maintenance

One of the advantageous features of this test set is simplification of maintenance. The *A*, *B*, *C*, and *D* relays (256 in all, or 1 set for each of the 64 terminals) are identically wired, and the OC, OCB, OCG, and GBC relays; (28 in all) are also identically wired. Thus, of a total of 384 relays in a test set, 320 fall in two groups of identically wired relays. This makes easier a good understanding of the circuit functions. A tape is provided for maintenance use in rapidly checking the entire test circuit. This tape not only places conditions on the test set to cause it to operate, but also places simulated defective conditions to make certain that the test set will stop and indicate, properly, such conditions.

Testing Instructions

A set of testing instructions(IMI), in addition to the perforated tape, is, provided by the engineer to the tester covering each different circuit to be tested. This instruction includes information for test set preparation, for action to be taken when the test set is stopped for a special condition and a complete number list of the tape perforations. Each line of perforations on a tape is represented by a 6-digit number. As explained heretofore, these numbers represent a phase of test set preparation determined by the tape group, *A*0 to *A*2, *B*0 to *B*7, *C*0 to *C*7, and sometimes, *D*0 to *D*7, in which they are located, plus the arrangement of the test set switching circuit. The test man, acquainted with the numbering significance and plan of perforations can, at a glance, determine how far the test has progressed, and what tests have gone on before. Referring to his chart Table I, which is quickly memorized, he can determine from the number list *D* tape group what conditions have been imposed since the last successful check line, as well as the unit terminal (*A*, *B*, and *C* tape groups) upon which the condition was imposed. At each check line is included identification of the relays that should be operated at that time. This information is given to aid in locating trouble conditions.

Conclusion

This test set is coming into extended use for testing wired relay units. The fact that lighted lamps indicate not only the nature of the trouble but pin-point the terminals to which the wires are connected that are in trouble, is of course, a great aid in finding the trouble.

◆

No Discussion

The Use of Electric Network Analyzers for Pipe Network Analysis

ROBERT E. STEPHENSON
ASSOCIATE MEMBER AIEE

J. R. EATON
MEMBER AIEE

THE determination of fluid-flow and pressure drop in complicated pipe networks such as gas or water distribution systems presents many problems similar to those encountered by the electrical engineer in the study of transmission and distribution networks. As the network analyzer has proved to be a very great help to the electrical engineer in his studies, it seems reasonable that similar techniques utilizing similar analyzers would be of considerable value in the analysis of pipe networks. Several investigators have reported methods for applying network analyzer techniques to gas or water distribution problems, some using special handling of conventional analyzers and others using special analyzers built solely for fluid-flow problems.

This paper presents a newly developed method for adapting conventional a-c or d-c network analyzers to pipe network problems. It has the advantage of requiring a minimum of slide rule computations and provides a degree of accuracy of results comparable to those attained in studies of electric systems made on the same analyzer.

Comparison of Fluid Flow and Electric Flow Networks

A comparison of the basic relations pertaining to fluid flow and to electric flow in networks demonstrates that the two systems are very similar in certain respects but quite different in one important aspect. It might be stated that Kirchhoff's laws apply in both systems. In a fluid flow network the algebraic sum of the pressure drops around any closed loop must be equal to zero, whereas the algebraic sum of the flow into any pipe junction must total zero. Because of these similarities of the two systems, it is logical that in studying pipe systems by means of electric networks, the pressure drop in the gas systems should be represented by voltage drop in the electric system while flow in the fluid system is represented by current in the electric system. When this relationship is adopted, the topology of the fluid system is maintained in the electric system which represents it.

The one important difference between the fluid and the electric systems is in the relationship between flow and pressure drop which, (for incompressible flow) is of the form given in equation 1

$$Q=K_gP^n \qquad (1)$$

Paper 53-410, recommended by the AIEE Computing Devices Committee and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Kansas City, Mo., November 2-6, 1953. Manuscript submitted February 16, 1953; made available for printing August 21, 1953.

ROBERT E. STEPHENSON is with the University of Utah, Salt Lake City, Utah; J. R. EATON is with Purdue University, Lafayette, Ind.

The authors acknowledge the valuable help and assistance given by J. B. Ward, H. W. Hale, and D. L. Johnson, all of Purdue University, in performing the experimental work connected with this development.

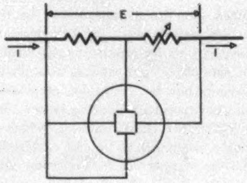


Fig. 1. Basic connections

where

Q = the flow through a particular pipe section
K_q = constant of proportionality
P = pressure drop or head-loss through a particular pipe section
n = a constant having the value 0.5 for gas flow and the value 0.54 for water flow

This equation is in contrast to the behavior of electric circuit elements in which current and voltage drop are related as indicated in equation 2

$$I=GE \tag{2}$$

where

I = the current through a particular element
E = the voltage drop across the element
G = the conductance of the element

In representing a pipe network by a topologically similar electric network, the first two characteristics of the fluid flow system are inherently maintained by the electric network. For an electric network to perform in a fashion identically analogous to the pipe network, it would be necessary to have current flow in each circuit element related to voltage drop across that circuit element by the relation

$$I=K_eE^n \tag{3}$$

Rewriting this equation as

$$I=(K_eE^{n-1})E$$

it is evident that this relationship may be attained if

$$G=K_eE^{n-1} \tag{4}$$

which states that the conductance of each circuit element of the electric network must change in value with changes in voltage drop across that section. The automatic adjustment to fit this equation requires elements whose conductance is a nonlinear function of the voltage impressed. On the other hand, manual adjustment of linear elements may be made to achieve the same result, provided a convenient method is available for indicating the proper adjustment.

Method of Adjusting *E-I* Relationship

When displayed on rectangular coordinates, a plot of the E versus I relationship of a linear conductance is a straight line through the origin, the slope being dependent on the value of the conductance. A similar plot of the E-I relationship where the conductance is varied according to equation 4 will be a parabola if n has a value of 0.5 (as for gas flow) or an approximate parabola if n has a value of 0.54 (as for water flow).

The relation between E and I for the variable conductance element can be exhibited on a cathode-ray oscillograph by producing a deflection along one axis proportional to current, and a deflection along the other axis proportional to voltage. If the conductance of the circuit element is adjusted either automatically or manually, in such a fashion that E and I follow the relation shown in equation 3, the figure shown on the cathode-ray oscillograph will be a section of a parabola with the zero signal position of the beam coincident with the origin. It follows, therefore, that if a parabola of proper shape is drawn on the face of a cathode-ray oscillograph, proper adjustment of a resistor in accordance with equation 3 will be indicated when the oscillograph beam coincides with the plotted line.

If, on the other hand, linear resistive elements are connected as shown in Fig. 1, and carry sinusoidal alternating current, the trace on the cathode-ray oscillograph will be a straight inclined line passing through the origin of the plotted line. The slope of this luminous straight line trace will be governed by the resistance (or conductance) of the element. The correct nonlinear relationship between current and voltage will be indicated when the end of this straight luminous trace coincides with the plotted line.

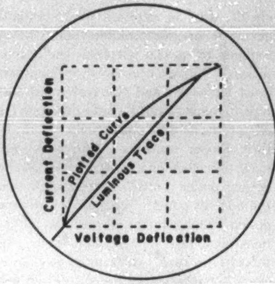


Fig. 2. View of the indicator screen

Proper adjustment of the resistance of the circuit element can be used to change the slope of the luminous trace and bring it into a position where the end of the trace coincides with the plotted line. This condition is illustrated in Fig. 2.

If the linear circuit element carries direct current, the beam will appear as a stationary spot, changing position only when voltage and current or circuit adjustment are changed.

The installation of an attenuator between the circuit element terminals and the oscillograph amplifiers makes possible the use of the same parabola on the oscillograph in the adjustment of different circuit elements having different values of K_e. Proper connection of the oscillograph to the network-analyzer metering circuits makes possible the transfer of the oscillograph and potentiometer combination to the various line elements of the network, the switching being accomplished by means of the conventional analyzer-metering selector system.

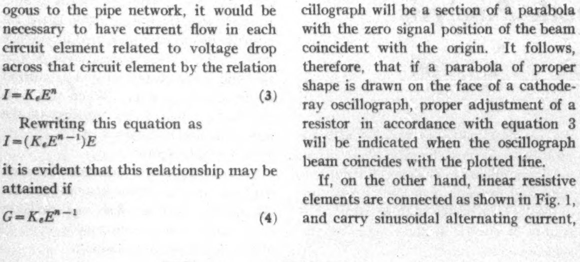


Fig. 3 (left). Switching device circuit diagram

S_1—Signal reversing switch
T_1 and T_2—Isolating transformers
S_2 and S_5—Ganged indicator range switches
S_4 and S_6—Zero position checking switches
Sh—Current shunt
S_3—Pipe potentiometer range selector
X—Pipe potentiometer
O—Oscilloscope
A_1 and A_2—Oscilloscope amplifiers
F, G, H, J—connections to network analyzer

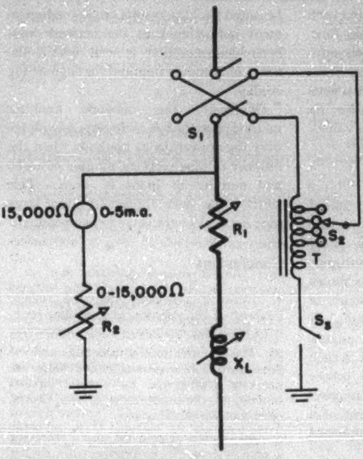

S₁
15,000Ω
0-5m.a.
R₁
S₂
T
0-15,000Ω
R₂
X_L
S₃

Fig. 4. Load adjuster

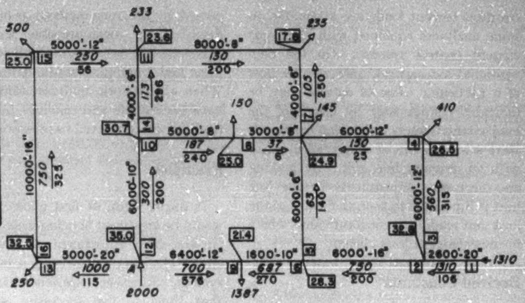

500
233
235
5000'-12"
8000'-8"
25.0
15
250
56
23.6
11
130
200
17.8
4000'-6"
113
286
4000'-6"
105
250
10000'-16"
750
325
150
145
410
30.7
14
5000'-8"
3000'-8"
6000'-12"
10
187
240
25.0
8
37
6
7
24.9
130
25
4
26.5
6000'-10"
300
200
6000'-6"
63
140
6000'-12"
560
315
32.5
35.0
21.4
32.8
16
12
3
5000'-20"
6400'-12"
1600'-10"
6000'-16"
2600'-20"
1310
13
1000
115
A
700
576
9
687
270
6
5
28.3
730
200
2
1310
106
1
250
2000
1387

Fig. 5. Solution by calculation

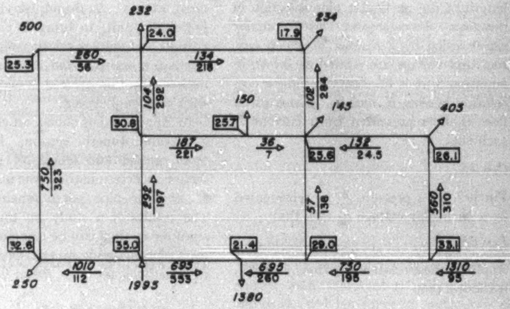

232
234
500
24.0
17.9
25.3
280
56
134
216
104
292
108
294
150
145
405
30.8
25.7
187
221
36
7
25.6
132
24.5
26.1
750
323
292
197
57
138
560
310
32.6
35.0
21.4
29.0
33.1
1010
112
695
333
695
260
730
195
1310
93
250
1995
1380

Fig. 6. Network analyzer solution

Circuitry

The success of the oscilloscope method of adjusting circuit elements is dependent on establishing a convenient method of switching the indicating circuit from one analyzer element to another, and conveniently adapting it to necessary changes to take into account the differences in pipe sections represented. A complete diagram of the indicating equipment for connection in proper relation to the analyzer metering circuits is shown in Fig. 3. Points F and G connect across the terminals of the analyzer master voltmeter, while terminals H and J are connected in series with the master ammeter. As a result, selector switching of the master instruments carries along the indicating equipment necessary for resistance adjustment.

As indicated in Fig. 3, the master instruments are isolated from the oscilloscope by means of transformers T_1 and T_2. It is possible that, when applying this method to some analyzers, better results would be obtained by replacing these transformers with isolating amplifiers. The particular analyzer to which this circuit was adapted is one in which currents may be as high as 3 amperes and voltages as high as 200 volts. As a result, the burden imposed by the indicating network was not of great importance but might be of considerable significance on analyzers utilizing smaller base voltages and currents.

To make the parabola drawn on the oscillograph representative of current and voltage relations in a pipe of any length and diameter, an appropriate attenuator, herein designated as a pipe potentiometer, is inserted between the isolating transformer and the terminals of the oscilloscope. This potentiometer must be of the linear precision type which may be set by dial indication to any desired ratio. Since the adjusting indicator may be used to adjust analyzer elements across which voltage may be of high or low value, and the current through it may be of high or low value, it is necessary to have in the circuit a means of changing the sensitivity of the instrument. This is accomplished by the gang-operated switches in both the voltage and current circuits. It should be noted that if the change of switch position in the current circuit reduces the vertical deflection of the oscilloscope trace to some fraction of its former value, say one-third, the corresponding change in the potential circuit must cause a reduction of deflection on the horizontal plates by a factor of $(1/3)1/n$.

The calibration of the completed oscilloscope circuit is accomplished by setting up a special test circuit on the analyzer and assuming for this circuit a particular value K_e. The current through this element, the voltage across it, and its resistance are then adjusted to values consistent with the assumed K_e. With an arbitrarily selected value of the setting of the pipe potentiometer, the gain controls of the cathode-ray oscilloscope are adjusted to bring the oscillograph spot in coincidence with the parabola. These oscillograph controls are then left undisturbed and the metering equipment switched to any other selected analyzer unit with the pipe potentiometer set to the appropriate value for that unit. Adjustment of the resistance of that particular unit may then be made to bring it into proper current and voltage relationship as indicated by the oscilloscope.

In conducting studies of fluid flow systems it is frequently necessary to represent loads of prescribed outflow. To represent such loads it is necessary to have appropriate circuits in the electric network through which current may be maintained at a constant value. These constant current loads are attainable on some anzlyzers equipped with electronic current control systems. On the conventional a-c network analyzer the flow of a particular value of current may be indicated continuously by means of the load circuit meters as shown in Fig. 4. Here a load circuit is connected in series with an ordinary line unit. By making any necessary adjustments in the line unit it is possible to keep the meter of the load unit reading at constant value, which insures constant load current.

Derived Constants

In equation 1, the constant K_g is indicated as being dependent on the size and length of the particular pipe selection in question. To represent this particular pipe section by a resistor in which current and voltage are related as shown in equation 3, it is necessary to find a correlation betwen K_g and K_e. Let a given flow Q_1 be represented by a current I_1 such that

$$Q_1 = AI_1$$

Further, let a pressure P_1 be represented by a particular voltage E_1 such that

$$P_1 = BE_1$$

Applying these relationships to equation 1, it follows that

$$K_e = \frac{K_g B^n}{A} \tag{5}$$

As mentioned in a preceding section, the pipe potentiometer is set to some arbitrary value, and calibration is made in a circuit element having a known value of K_e. If this known value is K_{e1} and the potentiometer setting is X_1, then the potentiometer setting X_n for any other value of K_{en} will be

$$X_n = X_1 \frac{(K_{e1})}{K_{en}} \tag{6}$$

Flow of Compressible Fluid

It is shown in the literature that the flow of compressible fluid through a pipe section may be represented by an equation similar to equation 1, but with P replaced by Pa where

$$Pa = \frac{P_1^2 - P_2^2}{2}$$

where P_1 is upstream pressure and P_2 is downstream pressure in absolute units. As it may be readily shown that the sum of the Pa's around a closed loop must equal zero, it is evident that the compressible fluid problem may be handled as far as the network analyzer is concerned, by the same methods as have been described for incompressible fluid flow. It is necessary only to handle all computations and manipulations in terms of Pa. When all network analyzer adjustments have been made and readings taken, Pa readings are converted to pressure values.

Examples

It might seem, at first glance, that it would be an almost hopeless task to make cut-and-dried adjustments of the many circuit elements representing a multipipe system. However, experience has shown that three to five successive series of consecutive manipulations are sufficient to bring the network into proper adjustment setting. As the adjustment of each individual unit, to bring it into proper current and voltage relationship, requires but a few seconds, the adjustment of a network by this method is accomplished with reasonable speed.

In Appendix I is shown an example of a 16-pipe 4-mesh system. From the cross section and length of each pipe section, a determination was made of K_g, K_e, and the pipe potentiometer setting. The results of a study by the network analyzer method can be compared with a similar study made by conventional slide rule methods, see Figs. 5 and 6. It may be noted that the results are in good agreement indicating that the network analyzer determination is well within the range of accuracy required for engineering studies.

Obviously, the network analyzer method of study has increasing advantage over the conventional methods when the number of pipe sections, number of loads, and number of loops is large. This method has been applied to rather large pipe networks with satisfactory results.

Conclusions

1. Network analyzers may be readily adapted to the solution of fluid flow problems involving complicated pipe networks.
2. The flow relations at junctions and the pressure relations around closed loops are properly represented by electric circuits because of the conformity of the electric circuit to Kirchhoff's laws.
3. An electric system set up to represent a pipe system will be topologically similar to the pipe system.
4. A method has been presented for quickly and easily relating the flow of current in each analyzer unit to the voltage across it so that current and voltage relations are similar to the flow and pressure drop relations applying to fluid flow in pipes.
5. The accuracy of the network analyzer method is equivalent to that obtained by slide rule calculations.

Table I. Measurements and Potentiometer Settings for Pipes

Pipe No.	Length	Diameter	K_g	K_e	X	Calculated Flow 1,000 Cubic Feet per Hour	Measured Flow 1,000 Cubic Feet per Hour
1	2,600	20	128,000	0.34	0.41	1,310	1,310
2	6,000	16	54,000	0.142	0.98	750	750
3	6,000	12	30,600	0.081	1.73	560	560
4	6,000	12	30,600	0.081	1.73	150	152
5	6,000	6	5,300	0.014	10.00	63	57
6	1,600	10	40,500	0.107	1.31	687	695
7	4,000	6	6,500	0.0172	8.13	105	102
8	3,000	8	15,100	0.040	3.50	37	36
9	6,400	12	29,000	0.078	1.80	700	695
10	5,000	8	12,100	0.032	4.38	187	187
11	8,000	8	9,100	0.024	5.83	130	134
12	6,000	10	21,200	0.056	2.50	300	292
13	5,000	20	121,000	0.24	0.58	1,000	1,010
14	4,000	6	6,500	0.0172	8.13	113	104
15	5,000	12	34,000	0.090	1.56	250	260
16	10,000	16	41,200	0.109	1.28	750	750

The potentiometer settings X were found using the relationship

$$X_n = X_a \frac{K_{ea}}{K_{en}}$$

The smallest K_{ea} was found and the corresponding X_a was chosen arbitrarily as being equal to 10. All other potentiometer settings X_n were then calculated.

After the network has been adjusted, the current and voltage distribution was measured. Since 1 ampere represented 1,000,000 cubic feet per hour, the current reading could be converted to thousands of cubic feet per hour by multiplying the current in amperes by 1,000. Likewise, since 100 volts represented a Pa of 700, the voltage drops could be multiplied by 7 to obtain the equivalent Pa. With the Pa thus obtained for each pipe, the actual pressure could be found at all points if it were known at one point. The gauge pressure was specified as 35 pounds per square inch at point A. All other pressures could then be calculated, since

$$Pa = \frac{(P_1 + P_2)}{2}(P_1 - P_2)$$

Knowing Pa from the preceding work, and calling $P_1 = (35+15) = 50$ pounds per square inch absolute P_2 could be found. The results of the network analyzer solution are shown in Fig. 6.

6. The time requirements for network analyzer solutions of fluid flow problems are favorable to the method, particularly in complicated pipe systems with many closed loops.

Appendix I. Analysis of an Intermediate Pressure Gas System

The complete solution of an intermediate pressure gas system by this newly developed method, as well as by conventional calculation, is given in the following material.

A map showing system layout, pipe lengths, pipe diameters, loads, supply pressures and pipe numbers is given in Fig. 5, along with the solution of pressure and flow distribution, as obtained by slide rule calculation. The slanted numerals above the flow arrows designate flows in thousands of cubic feet per hour. The vertical numerals under the flow arrows designate the *Pa* drop in each section. The gauge pressure is indicated at each junction by the numbered flag. The network analyzer solution is given in Fig. 6.

Table I shows the length, diameter, K_g, K_e, and X (potentiometer setting) for each pipe with flows obtained by each method.

K_g was obtained using a gas flow computing slide rule.

K_e was obtained using the relationship of equation 5.

In this case, 1,000,000 cubic feet of gas per hour corresponded to 1 ampere giving

$$A=\frac{10^6}{1}=10^6$$

Also 7 units of *Pa* corresponded to 1 volt, giving

$$B=\frac{7}{1}=7$$

Therefore

$$K_e=K_g\frac{\sqrt{7}}{10^6}=K_g\times10^{-6}\sqrt{7}$$

Based on these relationships, the current in amperes in any element would be interpreted as a flow in millions of cubic feet per hour in the corresponding pipe section. Likewise a voltage drop in volts would be interpreted as a *Pa* drop seven times as large numerically.

References

1. HYDRAULIC ANALYSIS OF WATER DISTRIBUTION SYSTEMS BY MEANS OF AN ELECTRIC NETWORKS ANALYZER, T. R. Camp, H. L. Hazen. *Journal*, New England Water Works Association, vol. 48, Dec. 1934, pp. 383 ff.

2. HYDRAULICS OF DISTRIBUTION SYSTEMS—SOME RECENT DEVELOPMENTS IN METHODS OF ANALYSIS, T. R. Camp. *Journal*, New England Water Works Association, Boston, Mass., vol. 57, no. 4, Dec. 1943, pp. 334–68.

3. DIRECT READING ELECTRIC ANALYZER FOR PIPELINE NETWORKS, M. S. McIlroy. *Journal*, American Water Works Association, New York, N. Y., vol. 42, no. 4, April 1950, pp. 347–66.

4. NETWORK FLOW ANALYSIS SPEEDED BY MODIFIED ELECTRICAL ANALOGY, H. A. Perry, D. E. Vierling, R. W. Kohler. *Engineering News-Record*, New York, N. Y., vol. 143, no. 12, Sept. 22, 1949, pp. 19–23.

5. SOLUTION OF HYDRAULIC DISTRIBUTION SYSTEM PROBLEMS UTILIZING A.C. NETWORK CALCULATOR, M. V. Suryaprakasam, G. W. Reid, J. C. Geyer. *Journal*, American Water Works Association, New York, N. Y., vol. 42, no. 12, Dec. 1940, pp. 1154–64.

6. GAS DISTRIBUTION PROBLEMS SOLVED BY ELECTRIC NETWORK CALCULATORS, J. P. Clennon, J. K. Dawson. *Bulletin DMC-51-14*, American Gas Association, New York, N. Y., April 1951.

No Discussion

An Analysis of an Analogue Solution Applied to the Heat Conduction Problem in a Cartridge Fuse

A. E. GUILE
NONMEMBER AIEE

E. B. CARNE
ASSOCIATE MEMBER AIEE

Synopsis: The general equation of heat conduction for a wire heated by an electric current and contained in an insulating cartridge filled with powder cannot be solved by analytical methods. In this paper, two approximate solutions are obtained; the first, obtained by an analytical method, assumes that radial heat loss from the wire is negligible; the second assumes that the parameters of the wire, filler, and cartridge remain constant at some average value, so that the equation can be solved by an analogue method. The analogue is described and the results obtained are checked against those calculated by assuming no radial heat loss, the circuit of the analogue being modified to eliminate the elements representing that loss. The magnitudes of the various errors associated with the analogue are determined as far as possible. The range of fusing current over which the analytical solution with no radial heat loss gives a good approximation to the correct temperature distribution is examined, and found to cover most of the region which is important in fuse operation. The temperature distributions obtained are used to determine the distribution of axial and radial losses along the length of a particular wire. From this the variation of losses with fusing current is obtained, and the range of current over which heat losses are negligible is determined.

When CURRENT flows in a conductor, energy is generated according to the well-known equation

$$\frac{dU}{dt}=i^2R \qquad (1)$$

where

U = energy generated
i = instantaneous current flowing in conductor
R = instantaneous resistance of conductor

If the current be assumed to be constant, and the resistance to vary with temperature θ in a linear manner, the rate of heat generation may be represented graphically, as in Fig. 1. Suppose that the rate of heat loss to the environment varies with temperature as shown by curve *A*, and let curve *B* and curve *C* represent the rates of heat generation appropriate to two currents, I_B and I_C $I_B>I_C$, then the conductor will remain in the solid state when carrying a current I_C, but will melt and break up if the magnitude be increased to I_B. The current at which the wire just melts is known as the minimum fusing current for the combination (wire and environment). Currents such as I_B will be referred to as fusing currents.

In studying fuse behavior, a knowledge of the distribution of temperature

---

Paper 53-411, recommended by the AIEE Computing Devices and Switchgear Committees and approved by the AIEE Committee on Technical Operations for presentation at the AIEE Fall General Meeting, Kansas City, Mo., November 2–6, 1953. Manuscript submitted March 10, 1953; made available for printing September 8, 1953.

A. E. GUILE is with the Queen Mary College, University of London, London, England, and E. B. CARNE is with Remington Rand, Inc., South Norwalk, Conn.

The work described was performed in the Electrical Engineering Department of Queen Mary College. The authors would like to acknowledge the advice and encouragement given by Prof. W. J. John, and by Dr. H. Tropper and W. B. Diamond. Mr. Carne would like to thank the Governors of Queen Mary College and the Senate of the University of London for financial support.

The section headed "Solution of Equation with Constant Parameters by Analogy" is a summary of part of a thesis presented by E. B. Carne to the University of London in fulfillment of the requirements for the degree of Doctor of Philosophy.

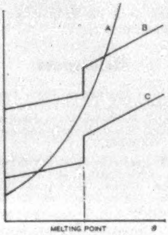


Fig. 1. **Definition of fusing current. Curve A represents the rate of heat loss from the wire to environment. Curves B and C are the rates of heat generation appropriate to two currents I_B and I_C, $I_B > I_C$. The conductor will remain in the solid state when carrying a current I_C, but will melt and break up if the current be I_B**

along the element is of great importance.[1] If the distribution has the form of A in Fig. 2, it is probable that the element will break at one point near the center of the wire. In a high-voltage a-c circuit, or in a lower voltage d-c circuit, this may lead to the formation of a stable, single arc, and the liberation of a considerable amount of energy with consequent unsatisfactory operation of the fuse, and unnecessary prolongation of the fault. On the other hand, if the temperature distribution is as shown in Fig. 2(B) a substantial length of the element melts and breaks up at the same time, and the fault current is ruptured with comparative ease. Because of the small diameter wires normally employed as fuse elements, and the fairly rapid heating which occurs, any attempt to determine the temperature distribution experimentally involves considerable difficulty. This paper presents a more general approach by approximate solutions of the general equation of heat conduction for the element.

The particular system to which these principles are applied was one which could be constructed easily in the laboratory from material at hand, and which was well within the range of the testing facilities available. Details of its construction are shown in Fig. 3, while the appropriate physical parameters for the materials used are given in Table I. The normal rated working current of such a fuse would be of the order of 10 amperes. If

k_0, k_1, k_2 = thermal conductivity of wire, filling powder, cartridge
c_0, c_1, c_2 = specific heat of wire, filling powder, cartridge
m_0, m_1, m_2 = density of wire, filling powder, cartridge
r_0 = radius of wire
r_1 = inner radius of cartridge
r_2 = outer radius of cartridge
A_0 = cross-sectional area of wire
K_0 = electric conductivity of wire
I_0 = current flowing through wire
(h_c+h_r) = combined coefficient of heat loss by convection and radiation from the outer surface of the cartridge[3]
$2L$ = total length of fuse element
J = Joule's equivalent
θ = excess temperature, ambient as datum
z = axial co-ordinate
t = time from commencement of current flow

then the equation of heat conduction for the system of Fig. 3 is

$$k_0A_0\frac{\partial^2\theta}{\partial z^2}+2\pi r_0k_1\left.\frac{\partial\theta}{\partial r}\right|_{r=r_0(+)}+\frac{I_0^2}{JA_0K_0}=A_0c_0m_0\frac{\partial\theta}{\partial t} \tag{2}$$

where the quantity

$$\left.\frac{\partial\theta}{\partial r}\right|_{r=r_0(+)}$$

is to be determined from

$$k\left[\frac{\partial^2\theta}{\partial z^2}+\frac{\partial^2\theta}{\partial r^2}+\frac{1}{r}\cdot\frac{\partial\theta}{\partial r}\right]=cm\frac{\partial\theta}{\partial t} \tag{3}$$

with conditions

(a) $r_0 \leqslant r \leqslant r_2$
(b) $k=k_1$ $c=c_1$ $m=m_1$ when $r_0<r<r_1$
(c) $k=k_2$ $c=c_2$ $m=m_2$ when $r_1<r<r_2$
(d) $\frac{\theta}{(h_c+h_r)}+k_2\left.\frac{\partial\theta}{\partial r}\right|_{r=r_2(-)}=0$
(e) $t=0$ $\theta=0$ for all z and r $-L<z<+L$
(f) $z=\pm L$ $\theta=0$ for all r and t

The initial and boundary conditions for equation 2 are

(g) $t=0$ $\theta=0$ for all z $-L<z<+L$
(h) $z=\pm L$ $\theta=0$ for all t

A general solution is not practicable. A solution for infinitely long wires has been given by Schubert,[3] but is not suitable for the investigation of temperature distribution, since it neglects axial heat flow which gives rise to the characteristic distributions shown in Fig. 2. Provided certain assumptions are made, however, two approximate solutions can be obtained, one analytical, the other from an analogue. The analytical solution involves the assumption of no radial heat loss and that the thermal parameters of the wire are constant. This neglect of radial heat loss is justifiable in the important practical case of very rapid heating, such as occurs on short circuit. The analogue is based on the assumption that the wire and filler parameters remain constant at some average value, as the temperature rises. The analytical method will be presented first and will be followed by a consideration of the analogue.

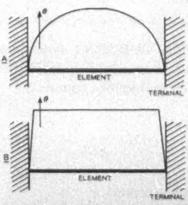


Fig. 2. **Idealized temperature distribution diagrams for fuse wires. If the distribution is as shown at A, the wire will probably break up at one point near the center; if the distribution is as shown at B, a considerable length of the element will melt and break up at the same time**

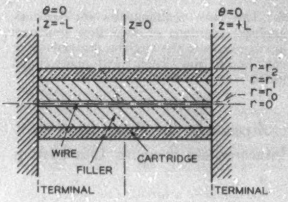


Fig. 3. **Details of the construction of the test fuse**

Solution of Equation Neglecting Radial Heat Flow

If radial heat losses can be neglected, and the thermal parameters of the wire may be treated as constant, equation 2 becomes

$$k_0A_0\frac{\partial^2\theta}{\partial z^2}+\frac{I_0^2}{JA_0}\rho_0(1+\alpha\theta)=A_0c_0m_0\frac{\partial\theta}{\partial t} \tag{4}$$

where

$\rho_0(1+\alpha\theta)$ = the resistivity of the wire, assumed to vary linearly with temperature

with conditions

(a) $t=0$, $\theta=0$ for all z $-L<z<+L$
(b) $z=\pm L$, $\theta=0$ for all t
(c) $z=0$, $\partial\theta/\partial z=0$ for all t

Rewriting equation 4 and putting

$$a^2=\frac{k_0}{m_0c_0};\quad q=\frac{I_0^2\rho_0}{Jr_0^4\pi^2m_0c_0}$$

gives

$$a^2\frac{\partial^2\theta}{\partial z^2}+q(1+\alpha\theta)=\frac{\partial\theta}{\partial t} \tag{5}$$

Using the transformation

$$\Theta(z, p)=p\int_0^\infty \epsilon^{-pt}\theta(z, t)dt$$

equation 5 becomes

Table I. Physical Parameters for Material in Test Fuse

Physical Constants	Porcelain	−120+140 Quartz	Copper
Density, grams per cm³	2.4	1.4	8.64 (average 20 C→1083 C)
Specific heat, calories per gram	1.8×10^{-1}	1.9×10^{-1}	1.1×10^{-1} (average)
Thermal conductivity, centimeter-gram-seconds units	2.5×10^{-3}	4.8×10^{-4}	8.58×10^{-1} (average)
Electrical conductivity, mhos			1.86×10^{-5} (average)
Electrical resistivity, at 20 C, ohm-cm			1.69×10^{-6}
(h_c+h_r) calories per second per C per cm²	2.95×10^{-4}		

Materials:

Porcelain cartridge
−120−140 quartz powder filler
No. 20 Standard Wire Gauge copper wire

Dimensions:

$r_0=1.74\times10^{-2}$ cm. (average 20 C→1083 C)
$r_1=1.75$ cm
$r_2=2.5$ cm
$2L=14$ cm

$$\frac{\partial^2\Theta}{\partial z^2}+(q\alpha-p)\frac{\Theta}{a^2}+\frac{q}{a^2}=0 \qquad (6)$$

This has the solution

$$\Theta=\frac{q}{(p-q\alpha)}-\frac{q\cosh\dfrac{z\sqrt{(p-q\alpha)}}{a}}{(p-q\alpha)\cosh\dfrac{L\sqrt{(p-q\alpha)}}{a}} \qquad (7)$$

Inverting with the aid of a table of transform-pairs and a few simple properties of the Laplace transformation gives the solution for θ as

$$\theta=\frac{4q}{\pi}\sum_{s=1}^{\infty}\frac{(-1)^s}{(2s-1)}\frac{\cos\dfrac{m_s z}{a}}{(q\alpha-m_s^2)}\times\left\{1-\epsilon^{(q\alpha-m_s^2)t}\right\} \qquad (8)$$

where

$$m_s=\frac{\pi a(2s-1)}{2L}$$

Solution of Equation with Constant Parameters by Analogy

THEORETICAL CIRCUIT AND EQUATIONS OF ANALOGUE

Consider the circuit shown in Fig. 4 where

$$R=R_0\Delta x;\ C=C_0\Delta x;\ i=i_g\Delta x;\ R_1=\frac{v\Delta y}{2\pi y\Delta x};$$
$$R_2=\frac{v\Delta x}{2\pi y\Delta y};\quad R_3=\frac{\eta 2\pi y_2}{\Delta x};\quad C_1=\gamma 2\pi y\Delta x\Delta y$$

and

$v=v_1,\quad \gamma=\gamma_1,\quad y_0<y<y_1$
$v=v_2,\quad \gamma=\gamma_2,\quad y_1<y<y_2$

The quantities, $R_0C_0i_g\eta v_1v_2\gamma_1$ and γ_2 are constants.

It can be shown that, in Limit ($\Delta x\rightarrow0$, $\Delta y\rightarrow0$) and provided $(i/y)(\Delta^2V/\Delta y)\rightarrow0$, the differential equation for this system is

$$\frac{1}{R_0}\frac{\partial^2V}{\partial x^2}+\frac{2\pi y_0}{v}\left.\frac{\partial V}{\partial y}\right|_{y=y_0(+)}+i_g=C_0\frac{\partial V}{\partial t} \qquad (9)$$

where V represents the voltage developed in the mesh, and the quantity

$$\left.\frac{\partial V}{\partial y}\right|_{y=y_0(+)}$$

is to be determined from

$$\frac{1}{v}\left[\frac{\partial^2V}{\partial x^2}+\frac{\partial^2V}{\partial y^2}+\frac{1}{y}\frac{\partial V}{\partial y}\right]=\gamma\frac{\partial V}{\partial t} \qquad (10)$$

which has as a boundary condition

(a) $\displaystyle\frac{V}{\eta 2\pi y_2}+\frac{2\pi y_2}{v}\left.\frac{\partial V}{\partial y}\right|_{y=y_2(+)}=0$

By suitable connections, the conditions

(b) $t=0$, $V=0$ for all x and y $-S\leq x\leq S$
(c) $x=\pm S$, $V=0$ for all y and t $y_0\leq y<\infty$

can be impressed on the system, so that the circuit of Fig. 4 can serve as an electrical analogy of the thermal system of the simple cartridge fuse already described, provided the electrical and thermal parameters of the fuse are not dependent on temperature.

TRANSFORMING SYSTEM PARAMETERS

In transforming values from one system to the other, effective use can be made of the principle of similarity to reduce the time scale of the electric system as compared with the thermal system. For the sake of generality, suppose the scales of length and time in the two systems to be connected by the relations

$$\xi x_E=x_T \qquad (11)$$

$$\mu t_E=t_T \qquad (12)$$

where the subscripts E and T denote the electric and thermal systems respectively. Further, suppose that the potential quantities be given by

$$\zeta P_E=P_T \qquad (13)$$

and, finally, suppose that each term of the thermal equations be multiplied by a fourth coefficient ω. By comparing terms it can be shown that

$$R_0=\frac{\xi^2}{\omega\zeta}\frac{1}{k_0A_0} \qquad (14)$$

$$C_0=\frac{\omega\zeta}{\mu}A_0c_0m_0 \qquad (15)$$

$$i_g=\frac{\omega I_0^2}{JA_0K_0} \qquad (16)$$

$$v=\frac{\xi^2}{\omega\zeta}\frac{1}{k} \qquad (17)$$

where $v=v_1$ when $k=k_1$
and $v=v_2$ when $k=k_2$

$$\gamma=\frac{\omega\zeta}{\mu}cm \qquad (18)$$

where $\gamma=\gamma_1$ when $c=c_1$, $m=m_1$
and $\gamma=\gamma_2$ when $c=c_2$, $m=m_2$

$$\eta=\frac{1}{\xi^2\zeta\omega(h_c+h_r)4\pi^2y_2^2} \qquad (19)$$

The coefficients $\xi\zeta\mu$ and ω can be defined within the limitations of equations 11

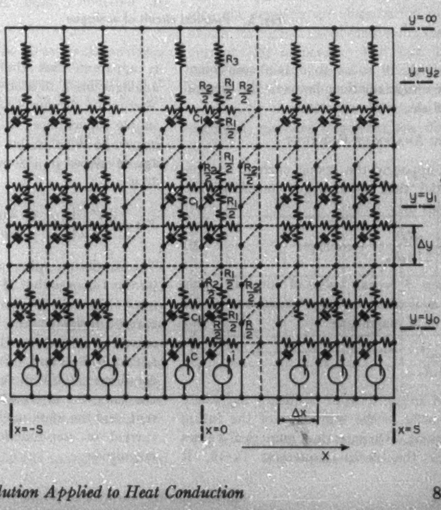


Fig. 4. Theoretical circuit of analogue

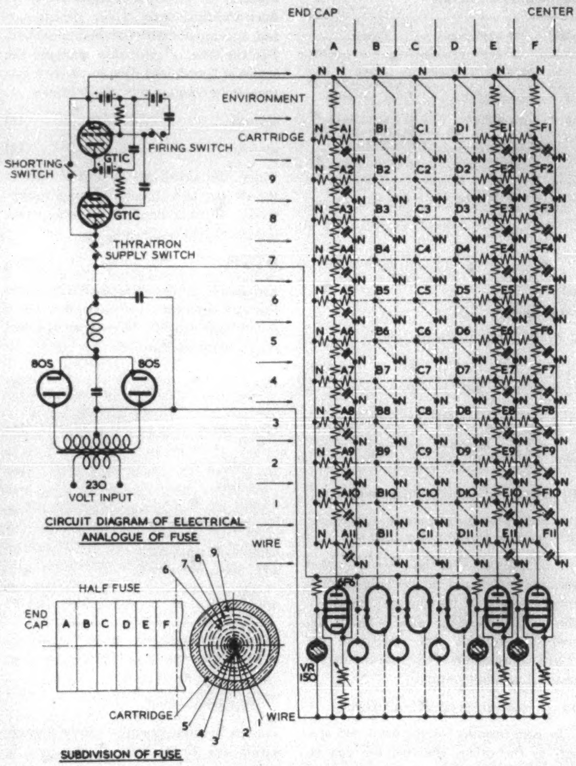


Fig. 5. Practical circuit of analogue

through 19 so as to obtain convenient magnitude relations between the thermal and electrical parameters.

THE ANALOGUE CURRENT

Supposing the fusing current is sinusoidal, then the analogue current will be given by

$$i_g=\frac{\omega I_p^2}{2JA_0K_0}(1-\cos 2\mu pt) \qquad (20)$$

where I_p = peak value of current. This is equivalent to two terms, one a d-c term of magnitude

$$\frac{\omega I_{rms}^2}{JA_0K_0}$$

and an a-c term of frequency $2\mu f$ where $f=p/2\pi$ is the frequency of the fusing current. On inserting numerical values into the design equations 14–19, it is apparent that the response of the analogue mesh to a 50 cycles per second alternating current is essentially that of a simple capacitance. Suppose this capacitance be denoted by C, then the voltage developed by the current of equation 20 is

$$V=\frac{\omega I_p^2}{2JA_0K_0C}\left[t_E-\frac{\sin 2\mu pt_E}{2\mu p}\right]$$

The maximum value of the alternating term is $1/2\mu p$ so that, if $t_E>>1/2\mu p$, i.e., $t_T>>1/2p$, only the d-c component of the current is significant. For 50 cycles per second fuse service this requires that $t_T>>1/628$ second. Provided that this condition is fulfilled, the response will be the same for alternating or direct current, and the mesh may be supplied with current of step-function shape and of magnitude

$$i_g=\frac{\omega I^2}{JA_0K_0}$$

where

I = actual magnitude of current (d-c)
= rms value of current (a-c)

A PRACTICAL CIRCUIT

The thermal fields are symmetrical about a cross-section normal to the longitudinal axis, and through the center of the fuse. The temperature on both sides of this plane will be the same, and there will be no heat flow across it. As far as the electrical system is concerned, this implies that the two halves may be separated one from the other, and the analogue circuit need only represent one half of the complete fuse.

The analogue circuits are true replicas of the thermal system only if they are characterized by continuously distributed parameters. In general, such circuits are difficult to build, and are easily disturbed by the introduction of measuring probes. Fortunately, approximate solutions can be obtained by the use of a finite number of lumped elements, the accuracy of the response increasing with the fineness of the subdivision. In the final assessment, this degree of subdivision is primarily dependent on considerations of economic cost and constructional time, and the analogue as constructed comprised 72 sections, the half-fuse in the electrical model being divided into six lateral discs and 12 longitudinal cylinders. A diagram of the analogue, and of the method of subdivision is given in Fig. 5.

The mesh is fed at six points by type *6F6* pentode valves working as constant current sources. Parallel stabilization of the screens is provided by type *VR150* glow-discharge tubes, and a step-function of current is generated by suddenly switching on the voltage to the anodes of the generators. Two *GT1C* thyratrons

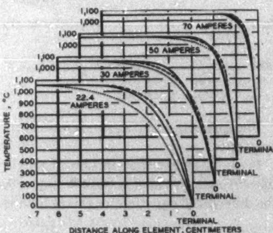


Fig. 6. A comparison of the temperature distributions obtained from the analytical solution and from the analogue, for various fusing currents

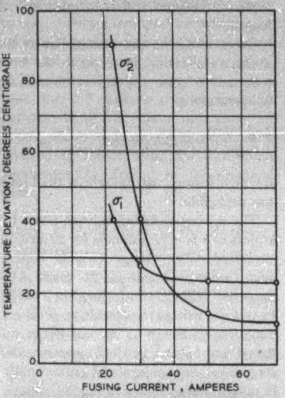


Fig. 7. The deviation between the analytical and analogue solution. σ_1 represents the deviation between the analytical system and the analogue modified so as to eliminate radial losses. σ_2 is the deviation between the analytical system and the complete analogue

are used in the switching circuit because it was found that one valve sometimes broke down before the triggering pulse was generated by closing the firing switch. The voltage rise at any mesh point is recorded by an oscilloscope,[4] the single-stroke time base being initiated by the applied signal, and a complete voltage (and therefore temperature) distribution diagram is built up by scanning each node in turn.

A number of tests were made on the analogue. The results of those significant to this paper are shown in Fig. 6.

Comparison of Analytical and Analogue Solutions

ERRORS INHERENT IN ANALOGUE

The response of the analogue will differ from that required to reproduce equation 2 because

1. The resistivity of the wire is represented by a constant average value.
2. The continuous thermal system is represented by a number of lumped circuits.
3. The analogue current magnitude has to be set up, and the tolerance on the values of analogue components kept as small as possible.
4. The analogue voltage response is measured by an oscilloscope.
5. The specific heat, thermal conductivity, and density of the wire, filling material and cartridge, and the coefficient of heat transfer from the outer surface of the cartridge to the ambient temperature are represented by average values.
6. The end terminals of the system are considered as infinite planes held at the ambient temperature.
7. In the case of an alternating fusing current, the fusing time must be very much greater than 0.0016 second.

COMPARISON OF ANALOGUE RESPONSE AND CALCULATED RESULTS TO ASSESS THE ACCURACY OF THE ANALOGUE

Temperature Distribution Errors in Analogue

The analogue was used to obtain temperature distribution curves with various fusing currents, for the case where radial heat flow was excluded. These curves are shown in Fig. 6 together with the curves obtained for the same currents by calculation from equation 8. Since both sets of curves represent the same case, a direct comparison between them can be made, any differences being due to errors in the analogue.

In this comparison, error 7 does not arise, since both solutions refer to a d-c fusing current; error 6 is common to both systems, and in any case is known experimentally to be small for terminals of appreciable size; error 5 is also common to both systems, and can be neglected in comparison with error 1, since the variation of specific heat and thermal conductivity of copper up to melting temperature is only of the order of 20 per cent, whereas the resistivity of copper increases by some 500 per cent over the same temperature range. The significant errors in this comparison are therefore 1, 2, 3, and 4.

The deviation (σ_1) between these two distributions has been calculated as follows: let

θ_1 = temperature rise given by the analogue at some point in the wire

θ_2 = temperature rise given by equation 8 at the same point in the wire for the same fusing current

and suppose the values of θ_1 and θ_2 are noted at n points along the wire length, then

$$\sigma_1=\sqrt{\left\{\frac{1}{n}\sum_{n}(\theta_1-\theta_2)^2\right\}}$$

i.e., the root mean square deviation in temperature. The magnitude of this deviation is plotted against fusing current in Fig. 7. It can be seen that, as the current decreases, the deviation increases from a constant value at high currents of approximately 20 degrees centigrade (C), i.e., 2 per cent of melting temperature, to about 50 C at the minimum fusing current.

Fig. 6 shows that the analogue gives too great a value of temperature rise near the terminals. This results from error 1 since, in the design of the analogue, a resistivity has been assumed for the fuse wire, which is the mean of the values at 20 C and 1,083 C. For high fuse currents, the greater part of the wire reaches melting temperature at the same time (e.g. 80 per cent for 70 amperes), and the use of this mean value along the whole length is justified. As the fusing time increases, the temperature distribution becomes less uniform and the use of this average resistivity produces too large a value of heat generation near the terminals so that the temperature there is too high.

The constant deviation of 20 C at high fusing currents suggests that, since error 1 becomes small in this region, this error of 2 per cent must be due to errors 2, 3, and 4 which are not dependent upon fusing current. Hence error 1 rises from zero at high currents to about 3 per cent at minimum fusing current, while the sum of errors 2, 3, and 4 remains constant at 2 per cent.

It will be assumed that this information on the errors can be used to correct the readings of the complete analogue when both radial and axial heat losses are included. The justifications for this are that the variation of the parameters of the filling powder and cartridge is very small and that the error resulting from the use of lumped circuits representing the filler will be much less than that for the wire, because of the greater number of circuits.

From these figures it therefore seems that the uncorrected analogue readings give the temperature distribution quite closely for currents above minimum fusing current (20 amperes). The fact that minimum fusing current is given as 20 amperes by the complete analogue and 5 amperes by the analogue with radial flow excluded, modifies somewhat the direct

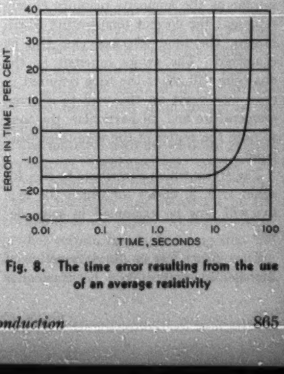


Fig. 8. The time error resulting from the use of an average resistivity

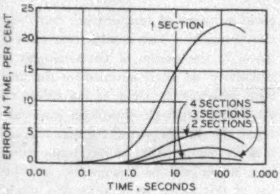


Fig. 9. The time error resulting from representing a continuously distributed system by a finite number of lumped sections

application of the deviations given by Fig. 7 to the complete analogue for very low currents. An upper current limit exists to the use of the analogue in the case of an alternating fusing current, since the thermal time must be very much greater than 0.0016 second so that heating effect of a step function of current in the analogue shall correspond to that of an alternating fusing current. For currents giving fusing times of this order, the temperature distribution will, however, be substantially rectangular, so that the range of the analogue is sufficient to cover those fusing currents in which the uneven temperature distribution will affect the breakup.

To determine temperature distributions for currents at or below the minimum fusing value, grading of the currents injected at the analogue wire nodes will take account of the variation of resistivity along the length, and reduce error 1. Comparison of the rise in temperature of an actual wire at a current just above the minimum fusing current as measured by the over-all increase in resistance shows good agreement with the resistance increase predicted by the analogue, until oxidation of the wire becomes appreciable just before the element melts.

Time Response Errors of the Analogue

Although the main purpose in constructing the analogue has been achieved insofar as the analogue readings closely represent the correct temperature distribution, the time response can also be examined. The errors inherent in the analogue will affect the time required to raise any point on the wire to a given temperature and, in particular, the time to raise the center of the wire to melting point.

The errors which contribute to this are again 1, 2, 3, and 4, of which the first two will now be examined in some detail.

Time Error Due to the Assumption of Average Resistivity—Error 1. If in equation 4, $\rho_0(1+\alpha\theta)$ is replaced by ρ_{av} (the average resistivity of solid copper between 20 C and 1083 C) as was done in the analogue, then instead of equation 8, the expression for the temperature rise is

$$\theta=\frac{4q_1}{\pi}\sum_{s=1}^{\infty}\frac{(-1)^s}{(2s-1)}\cos\left(\frac{m_s z}{a}\right)\times \frac{1}{m_s^2}\{\epsilon^{-m_s^2 u}-1\} \qquad (21)$$

where

$$q_1=\frac{I_0^2\rho_{av}}{Jr_0^4\pi^2 c_0 m_0}$$

If the time taken for the center of the wire to reach melting temperature given by equation 21 is compared with that given by equation 8 for the same current, then any difference is due to the assumption of average resistivity. This time error is shown plotted against time in Fig. 8. Until the minimum fusing current is approached, this error is constant at −15.5 per cent. The reason for this can be seen by considering a copper wire heated through a temperature rise of θ C with no heat losses, for the two cases (a) and (b)

(a) $\rho=\rho_{av}$
(b) $\rho=\rho_0(1+\alpha\theta)$

In case (a)

$$\frac{I_0^2\rho_{av}L}{A_0}\,dt=c_0A_0L\,d\theta$$

from which

$$\int_0^t\frac{I_0^2}{A_0^2}\,dt=\frac{c_0\theta}{\rho_{av}}=7.86\times10^8 \qquad (22)$$

when $\theta=1063$ C.

In case (b)

$$\frac{I_0^2\rho_0(1+\alpha\theta)L}{A_0}\,dt=c_0A_0L\,d\theta$$

from which

$$\int_0^t\frac{I_0^2}{A_0^2}\,dt=\frac{c_0}{\rho_0\alpha}\log_\epsilon(1+\alpha\theta)=9.30\times10^8 \qquad (23)$$

when $\theta=1063$ C.

$\int_0^t\frac{I_0^2}{A_0^2}\,dt$ is the current heat integral given by Gibson[5] and is independent of length.

Thus, owing to the assumption of an average resistivity, the error in the time required to raise the center to melting temperature is

$$\frac{7.86-9.30}{9.30}=-15.5\text{ per cent}$$

for high fusing currents where the losses are negligible.

The minimum fusing current given by equation 8 is 4.11 amperes for 29 British Standard Wire Gauge (SWG) copper wire, whereas equation 21 gives 5.05 amperes. (Since both neglect radial heat loss, neither represents a practical condition.) Hence as the current is reduced and approaches 5.05 amperes, the error must change sign and grow rapidly in magnitude as shown in Fig. 8.

Time Error due to Representing the Fuse by a Finite Number of Analogue Sections—Error 2. This error has been examined by representing the half length of fuse by circuits having one, two, three and four sections, and analyzing their transient responses. The times given by these circuits for various currents to raise the center of the wire to melting temperature, were compared with the times given by equation 21. An average resistivity is assumed in both cases so that any time error is due to the finite number of analogue sections. Fig. 9 shows how this time error varies with the fusing time and the number of sections. It can be seen that the analogue with its six sections does not have an appreciable time error resulting from this in any part of the fusing current range.

The time range in which the analogue was used was 0.08 to 1.2 seconds and in this range the errors result from:

1. Assumption of an average resistivity.
3. Setting up the analogue.
4. Measurement.

The error attributable to the first assumption amounts to −15.5 per cent in this range. The errors recorded by the analogue in excess of this are given in Table II; they must result from the inaccuracy of setting up and measurement.

Table II. Errors in Excess of Those Attributable to Assumption of Average Resistivity

Fuse Current, Amperes	Calculated Time, Seconds	Analogue Time, Seconds	Analogue Time Error, Seconds	Time Error, Per Cent
22.44	1.63	1.67	+0.04	2.45
30	0.912	0.88	−0.032	3.5
40	0.513	0.54	+0.027	5.25
50	0.328	0.33	+0.002	0.6
60	0.228	0.24	+0.012	5.23

The complete analogue predicts times to reach melting temperature which show very good agreement with those recorded in actual tests for times of the order of, or less than, 1 second. For times greater than this, the analogue predicts times 30 per cent to 40 per cent greater than the actual prearcing times. This is probably because of oxidation of the wire which accelerates the breakup in the practical case.

Range of Application of the Analytical Solution

In many of our problems it has been more convenient to calculate the temperature distribution than to set up the analogue, and it is therefore necessary to know over what range the approximation of no radial heat flow gives reasonable results, i.e., over what range equation 8 can be used.

This can be shown by comparison between the temperature distributions obtained from equation 8 and those given by the complete analogue, when the latter are corrected for the errors discussed. The rms deviation (σ_2) in temperature between the distributions obtained from the complete analogue and those calculated from equation 8 is plotted against fusing current in Fig. 7. It will be seen that above about 1.5 times minimum fusing current, the deviation is less than 4 per cent of melting temperature. The analytical solution therefore gives a good approximation to the correct temperature distribution for fuse currents above about 1.5 times minimum fusing current. As will be seen later, however, the times for the center to reach melting temperature predicted by this equation, are true only for fusing times smaller than a few hundredths of a second, where radial loss is indeed negligible.

Analysis of Axial and Radial Heat Losses in a Simple Cartridge Fuse

In any element of the fuse wire, the heat generated is expended in raising the temperature of the wire and supplying the radial and axial heat losses.

If there are no losses, equation 23 holds. This gives the value of the current-heat integral as 9.30×10^8 for a copper element at melting temperature. Consider the case of a fusing current of 70 amperes. The time required to raise the wire center to melting temperature calculated from equation 8 is 0.1677 second, and

$$\int_0^t \frac{I_0^2}{A_0^2}\,dt=\frac{70^2\times0.1677}{9.4^2\times10^{-8}}=9.30\times10^8$$

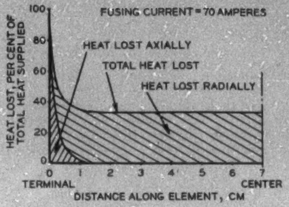


Fig. 10. The distribution of axial and radial losses along the fuse wire for a fusing current of 70 amperes

There is therefore no axial heat loss from an element at the center of the wire. Inspection of the temperature distribution with no radial loss in Fig. 6 shows constant temperature over a distance of about 5.5 cm from the center. Consequently there can be no axial heat loss in this region. Fig 6 also shows that the introduction of radial loss modifies the temperature distribution in such a way that this region decreases to 5 cm from the center. The time required for the center of the wire to reach melting temperature is 0.25 second. Hence, assuming no radial loss, 70 amperes raises 5 cm to melting temperature in 0.1677 second; with radial loss, 70 amperes raises 5 cm to melting temperature in 0.25 second, so that, in this region

$$\frac{\text{radial energy lost}}{\text{energy stored in wire}+\text{radial energy lost}}=\frac{0.25-0.1677}{0.25}=33\text{ per cent}$$

Again Fig. 6 shows that within 0.25 cm of the terminals the temperature distribution given by the analogue is unchanged by the addition of radial flow. At 0.2 cm from the terminals, 70 amperes raises the temperature by 440 C in 0.1677 seconds when radial flow is neglected. Equation 23 shows that if there is no loss 70 amperes raises the temperature by 440 C in 0.104 second.

Hence

$$\frac{\text{axial energy lost}}{\text{energy stored in wire}+\text{axial energy lost}}=\frac{0.1677-0.104}{0.1677}=38\text{ per cent}$$

At the fuse terminal, the axial heat loss will account for 100 per cent of the heat supplied. From calculations such as these, it is possible to estimate the variation of axial heat loss along the wire. The distribution is shown in Fig. 10.

When radial loss is included, 70 amperes raise the temperature of a point 0.2 cm from the terminals by 440 C in 0.25 second; therefore

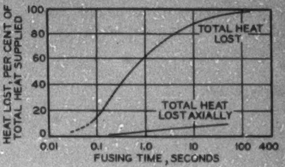


Fig. 11. The variation of the total heat lost with fusing time

$$\frac{\text{total energy lost}}{\text{energy stored in wire}+\text{total energy lost}}=\frac{0.25-0.104}{0.25}=58.4\text{ per cent}$$

The variation of the total heat lost along the wire is plotted in Fig. 10. The difference between this curve and the curve representing the distribution of axial heat loss gives the proportion of the heat lost by radial flow.

If this process is repeated for various fusing currents, and the mean values of the heat losses for the whole fuse element, expressed as a percentage of the total heat input, are determined, a curve of the variation of the losses against fusing time may be drawn. This is done in Fig. 11. The curve shows a total loss of 15 per cent at a heating time of 0.1 second, and agrees with the suggestion[6] that heat losses are negligible below 0.016 second.

Conclusions

The analogue described is capable of predicting the temperature distribution along the element of a cartridge fuse with a fair degree of accuracy over a range of fusing currents extending from minimum fusing current up to about five times minimum fusing current, i.e., about ten times the normal rated current.

For currents above about 1.5 times minimum fusing current, the temperature distribution may be calculated to a fair degree of accuracy on the basis of no radial heat loss. The fusing times given by this calculation are incorrect however except for very short fusing times where the radial loss is in fact negligible: for the fuse considered the time must be less than a few hundredths of a second for the radial and axial losses to be negligible.

The methods used are of quite general application, although the specific remarks made apply only to the particular size of cartridge fuse described at the beginning of the paper. It is possible to predict certain trends, knowing that the axial loss depends on the square of the wire diameter, and that the radial loss depends principally upon the diameter alone, but these should not be pressed to extremes.

References

1. A Mechanism for the Fuse Pre-Arcing Period, **E. B. Carne.** *AIEE Transactions*, vol. 72, Aug. 1953, pp. 593–99.

2. Heat Transmission (book), **W. H. McAdams.** McGraw-Hill Book Company, Inc., New York, N. Y., 1933.

3. On the Solution of the Heat Conduction Equation for an Electric Fuse (in German), **G. U. Schubert.** *Zeitschrift fur Angewandte Physik*, Berlin, Germany, vol. 2, 1950, pp. 174–78.

4. Recording Voltage Transients During Fuse Operation, **E. B. Carne.** *Engineer*, London, England, vol. 192, 1951, pp. 553–54.

5. The High Rupturing Capacity Cartridge Fuse, **J. W. Gibson.** *Journal*, Institution of Electrical Engineers, London, England, vol. 88, pt. II, 1941, pp. 2–40.

6. Performance Criteria for Current-Limiting Power Fuses—II, **E. W. Boehne.** *AIEE Transactions*, vol. 65, 1946, pp. 1034–44.

◆

No Discussion

LATE DISCUSSION

The following closure was received too late to be included with the paper.

Sensitivity and Output Formulas for the Resistance Bridge

Author's closure of paper 53-1 by Philip M. Andress, published in Communication and Electronics, May 1953, pp. 213–16.

Philip M. Andress: Dr. Harris and Dr. Wenner have mentioned in their discussion that it may be more convenient to consider galvanometer sensitivity in terms of available electromotive force and available current rather than current in the coil. This is certainly true in the calculation of steady-state deflections where the damping or the time response for specified precision of reading is to be considered. In the case of a dynamic or transient state, the ideas may have even greater significance affecting the design as well as the circuit application of the galvanometer.

To illustrate one useful application of the idea, we may consider the following statement: To reduce the response of a galvanometer by a given ratio without changing the damping, both the available electromotive force and the available current must be reduced by this ratio.

For example, assume it is desired to reduce sensitivity by 10 to 1 where the source resistance is 1,000 ohms. There are two possible L-sections. In the first, the series resistance is toward the source and serves to reduce the available current. This calls for 9,000 ohms since it raises the resistance looking toward the source by a factor of 10 to 1. According to the rule, it is then necessary to reduce the electromotive force by the same ratio and this is done by shunting to restore the original resistance of 1,000 ohms. The required parallel resistance to accomplish this is 1,111 ohms.

In the second method, the available electromotive force is reduced first by shunting to 100 ohms. This calls for a parallel resistance of 111.1 ohms. According to the rule, the available current must also then be reduced by the 10-to-1 factor. Hence the L is completed by adding a series resistance of 900 ohms on the galvanometer side.

It will be noted that a considerable amount of juggling of algebraic expressions is eliminated by this method, also that the calculation is entirely independent of the galvanometer characteristics. The deflection is reduced precisely by 10 to 1, yet there is no change in damping.

Addendum: The particular formulas for w, y, and z given in Table I of the paper, Auxiliary Equations 6, 7, and 8, apply where the input branch connects the junction of arms X and A with the junction of arms B and R.

Errantum: The title of Table I was listed incorrectly. It should read "Bridge Sensitivity and Output Formulas."

Under**, item 2 of the footnotes for Table I, delete $+A$ from the expression for z. If $L=\infty$, $z=[X+RC/(C+G)]/T$.